

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

**Cryo-EM analysis of mammalian
co-translational membrane protein insertion
and N-linked glycosylation**

Katharina Isabelle Braunger

aus

Ulm, Deutschland

2018

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Roland Beckmann betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 15.11.2018

Katharina Isabelle Braunger

Dissertation eingereicht am: 15.11.2018

1. Gutachter: Prof. Dr. Roland Beckmann

2. Gutachter: Prof. Dr. Karl-Peter Hopfner

Mündliche Prüfung am: 12.12.2018

Summary

In eukaryotes, membrane protein insertion and asparagine-linked glycosylation (N-glycosylation) occur co-translationally at the membrane of the endoplasmic reticulum (ER) for the vast majority of substrates. Both events are of critical significance for membrane protein topogenesis, trafficking of proteins to their final intra- or extracellular destination and protein functionality. Insertion is performed by the heterotrimeric Sec61 complex for most substrates while glycan transfer is catalyzed by the oligosaccharyltransferase (OST) - a complex of at least eight protein subunits which exists in two different isoforms (STT3A- and STT3B-OST) in higher eukaryotes. The transient and highly dynamic nature of the resulting ribosome-translocon complexes has precluded them from thorough structural characterization so far.

This thesis describes an experimental strategy for the *in vitro* generation and isolation of mammalian co-translational membrane insertion/N-glycosylation intermediates. The routine was used in order to obtain solubilized intermediates of the G-protein coupled receptor bovine opsin at defined stages of biogenesis for structural analysis by cryo-electron microscopy (cryo-EM). Preliminary structures of three different intermediates are presented and provide unexpected findings regarding the functional state of Sec61. It was observed in a so-called 'primed' rather than an 'inserting' conformation. Besides, the insertion/N-glycosylation substrate could not be located within the membrane region. Extended analysis of a very early intermediate led to the first high-resolution structure of a mammalian OST-containing ribosome-translocon complex. It revealed the spatial arrangement of mammalian OST subunits in the ER membrane. In addition, it explained how only one of the two naturally occurring paralogs of OST can associate with the ribosome and the membrane insertion machinery. Besides, the structure enabled building of a molecular model for the catalytic STT3A subunit and the paralog specific DC2 subunit. Comparison with recent structures of yeast OST confirmed a conserved complex architecture but also indicates potential differences among eukaryotic OST complexes. Moreover, the structural analysis uncovered an unassigned density segment at the Sec61-OST interface which was tentatively identified as the transmembrane domain of the *in vitro* insertion/N-glycosylation substrate.

Contents

1	Introduction	1
1.1	Ribosome architecture	1
1.2	Co-translational membrane insertion at the ER	2
1.2.1	Architecture of the Sec61 protein conducting channel	3
1.2.2	Biological functions of Sec61	4
1.2.3	Accessory translocon factors	6
1.2.4	Membrane protein terminology	8
1.2.5	Sec61-dependent membrane protein insertion - a thermodynamic view	9
1.2.6	TMH topology determinants	10
1.3	Biogenesis of bovine opsin	12
1.4	Asparagine-linked glycosylation at the ER membrane	14
1.4.1	Evolution and significance of OST	14
1.4.2	The oligosaccharide precursor	16
1.4.3	Catalytic mechanism of STT3-type enzymes	16
1.5	The eukaryotic OST complex	20
1.5.1	Complex architecture and distance restraints	20
1.5.2	The catalytic STT3 subunit	21
1.5.3	Shared accessory subunits	22
1.5.4	The STT3A specific subunits DC2 and KCP2	24
1.5.5	The STT3B specific oxidoreductase subunit	25
1.6	Aims of this thesis	25
2	Materials and Methods	27
2.1	General	27
2.2	Vectors	27
2.3	Molecular Cloning	27
2.3.1	Polymerase chain reaction (PCR)	27
2.3.2	Degradation of parental DNA	29
2.3.3	Phosphorylation and Ligation	30

2.3.4	Plasmid transformation	30
2.3.5	Plasmid isolation	30
2.4	Preparation of mRNA	30
2.4.1	Template generation by PCR	30
2.4.2	<i>In vitro</i> transcription	31
2.5	<i>In vitro</i> translation reactions	32
2.6	Protein analysis	34
2.6.1	Protein precipitation	34
2.6.2	Deglycosylation by Endoglycosidase H (Endo H)	34
2.6.3	SDS-Polyacrylamide-Gel-Electrophoresis (SDS-PAGE)	35
2.6.4	Semi-dry western blotting	35
2.6.5	Mass-spectrometry	36
2.7	Purification of opsin biogenesis intermediates for cryo-EM	37
2.8	Cryo-EM data collection	39
2.9	Cryo-EM data analysis and model building	40
3	Results	43
3.1	Expression of C-terminally truncated variants of bovine opsin	43
3.2	Preparation of solubilized opsin biogenesis intermediates	47
3.3	Preliminary cryo-EM structures of opsin biogenesis intermediates	49
3.4	Cryo-EM analysis of an early opsin biogenesis intermediate	52
3.5	Conformation of Sec61	55
3.6	Characterizing mammalian ribosome-bound OST	56
3.6.1	Evaluating OST-isoform specificity	56
3.6.2	Positioning of OST subunits	58
3.7	Interactions of OST with the ribosome and Sec61	61
3.7.1	The OST subunit DC2 at the Sec61 interface	61
3.7.2	The RPN1-Ribosome interface	62
3.7.3	Conformational dynamics in different functional states	63
3.8	Tentative positioning of the nascent insertion-glycosylation substrate	64
4	Discussion	69
4.1	A model for OST isoform specificity	69
4.2	Comparison of eukaryotic OST complexes	73
4.3	Opsin membrane insertion takes unexpected ways	79
5	Future perspectives	83

Appendix	85
A MS analysis of the solubilized OP96 biogenesis intermediate	87
A.1 Hits from the <i>canis lupus familiaris</i> database	87
A.2 Hits from the <i>oryctolagus cuniculus</i> database	96
References	109

List of Figures

1.1	Architecture of the heterotrimeric Sec61 channel.	3
1.2	Types of membrane proteins.	8
1.3	Topology and structure of bovine rhodopsin.	12
1.4	Bovine opsin biogenesis.	14
1.5	Composition of eukaryotic OST complexes	15
1.6	The LLO precursor.	16
1.7	CET structure of mammalian OST-containing RTCs.	21
3.1	Construct design and test expression of OP96 in absence or presence of PKRM.	44
3.2	Optimizing PKRM for use in <i>in vitro</i> translation.	46
3.3	Test expression of different opsin NC lengths in absence or presence of PKRM.	47
3.4	Cryo-EM sample preparation.	48
3.5	Purification of solubilized opsin biogenesis intermediates for cryo-EM.	49
3.6	Preliminary cryo-EM structures of thee opsin biogenesis intermediates.	50
3.7	Sec61 conformation in the solubilized opsin biogenesis intermediates.	51
3.8	Classification scheme of the combined cryo-EM datasets from the sol- ubilized OP96 biogenesis intermediate.	53
3.9	Local resolution distribution.	54
3.10	Cryo-EM structure of the programmed solubilized RTC.	55
3.11	Atomic model for Sec61.	56
3.12	CET of native RTCs.	58
3.13	Atomic model of mammalian STT3A.	59
3.14	Structural conservation of STT3-type proteins.	60
3.15	Spatial arrangement of mammalian OST subunits.	61
3.16	The Sec61-DC2 interface.	62
3.17	The ribosome-RPN1 interface.	62
3.18	Conformational states of different Sec61-OST complexes.	63
3.19	Additional density at the Sec61-OST interface.	65

3.20	Putative placement of the NC TMH.	65
3.21	OP96 NC connectivity.	66
4.1	Primary sequence alignment for the cytosolic STT3-TMH12/13 loop in various metazoan organisms.	72
4.2	Model for OST isoform specificity.	73
4.3	Conserved architecture of eukaryotic OST complexes.	74
4.4	Conformation of TMHs in STT3-type proteins.	76
4.5	Sequence conservation between different eukaryotic OST subunits. .	78
4.6	Superposition of a model for fungal OST and the cryo-EM density of mammalian STT3A-OST.	79
4.7	Expected architecture of opsin biogenesis intermediates.	80

List of Tables

1.1	Functionally important residues in STT3-type enzymes of different organisms.	18
2.1	Primers used for plasmid modification.	28
2.2	PCR program used for plasmid mutagenesis and amplification. . . .	29
2.3	PCR program used amplification of opsin constructs.	31
2.4	Reaction mix for <i>in vitro</i> transcription.	32
2.5	Reaction mix for <i>in vitro</i> translation without membranes.	33
2.6	Reaction mix for <i>in vitro</i> translation in the presence of PKRM.	34
3.1	Abundance of OST subunits in the solubilized OP96 biogenesis intermediate.	57
4.1	Conservation of primary sequences between STT3A and STT3B from various metazoan organisms.	71
A.1	MS analysis of the solubilized OP96 sample - Hits from <i>canis lupus familiaris</i> (dog).	87
A.2	MS analysis of the solubilized OP96 sample - Hits from <i>oryctolagus cuniculus</i> (rabbit)	96

Chapter 1

Introduction

1.1 Ribosome architecture

Ribosomes are the molecular machines which are responsible for protein synthesis. The molecular dimensions of these complexes are often referred to in Svedberg units (S), according to their sedimentation coefficient. A eukaryotic 80S ribosome consists of a 60S large ribosomal subunit (LSU) and a 40S small ribosomal subunit (SSU). They are composed of four different ribosomal RNA (rRNA) and approximately eighty ribosomal protein species. Ribosomal proteins and rRNA assemble in a highly complex and regulated way to form different functional sites. The LSU harbors the catalytic peptidyl transferase center (PTC) where individual amino-acids are covalently linked to the growing nascent chain (NC) via an ester bond. Furthermore, it contains the exit tunnel for the nascent polypeptide. The SSU comprises the messenger RNA (mRNA) tunnel and the decoding center (DC) where triplets of mRNA codons are sequentially presented to transfer RNA (tRNA) molecules. tRNAs are the crucial players for converting genetically encoded information from a nucleic acid sequence into a polypeptide chain format. They bind the ribosome at defined sites at the intersubunit space, recognizing the mRNA codon on the SSU and delivering the corresponding amino acid (aa) to the PTC on the LSU. The ribosomal tRNA binding positions are referred to as aminoacyl-, peptidyl- and exit sites (A-, P- and E- site, respectively). (For reviews see for example Melnikov et al., 2012; Wilson and Cate, 2012)

The structure of ribosomes has been studied extensively (Reviewed in Schmeing and Ramakrishnan, 2009; Voorhees and Ramakrishnan, 2013; Yusupova and Yusupov, 2014). Initial high resolution information was derived from crystal structures of the bacterial counterparts (Ban et al., 2000; Wimberly et al., 2000). Atomic-resolution models of the eukaryotic translation machineries have been described more recently by several x-ray crystallography and cryo-electron microscopy (cryo-EM) groups

(Ben-Shem et al., 2011; Anger et al., 2013; Voorhees et al., 2014; Khatter et al., 2015). The translation cycle in pro- and eukaryotes can be separated into four main stages: initiation, elongation, termination and recycling. The basic principles of translation and also many of the protein factors involved are highly conserved. However the eukaryotic cycle has adapted additional components and evolved to greater complexity.

Approximately 30 aa of an elongating NC are protected by the peptide exit tunnel of the LSU (Voss et al., 2006). A plethora of factors is known to bind the ribosome in proximity to the tunnel exit in order to act on the newly synthesized polypeptide as soon as it emerges from the ribosome (Kramer et al., 2009). In eukaryotes, the ribosomal proteins uL23/uL29 and uL22/eL31 which surround the tunnel exit have been identified as universal adaptor sites for ribosome-associated factors (Kramer et al., 2009; Pech et al., 2010; Ban et al., 2014). These factors include chaperones and NC modifying enzymes such as methionine aminopeptidases or N-acetyltransferases. In addition, this region is the binding platform for several components of the early secretory pathway, such as the signal recognition particle (SRP), the signal recognition particle receptor (SR) or the heterotrimeric Sec61 complex.

1.2 Co-translational membrane insertion at the ER

In eukaryotes 20-30 % of all open reading frames are predicted to encode α -helical integral membrane proteins (Von Heijne, 2011). The majority of these proteins are integrated into the lipid bilayer co-translationally (Rothman and Lodish, 1977) during the early steps of the secretory pathway. The pathway initiates when a ribosome engaged in translation of a secretory or designated transmembrane protein (TMP) is recognized by SRP and targeted to the SR in the endoplasmic reticulum (ER) membrane (reviewed in Akopian et al., 2013; Nyathi et al., 2013). NC recognition occurs via SRP binding to a hydrophobic signal sequence (SS) in the translated peptide. GTP hydrolysis by SRP and SR facilitates handover of the ribosome-nascent chain complex (RNC) to the protein conducting channel (PCC) in the ER membrane. The PCC is conserved across all three kingdoms of life and can translocate nascent polypeptide stretches across or insert them into the membrane. In eukaryotes, the PCC is formed by the heterotrimeric Sec61 complex and is located in the ER membrane. The homologous SecYEG complex in bacteria (SecYE β in archaea) is embedded in the inner plasma membrane.

1.2.1 Architecture of the Sec61 protein conducting channel

First direct evidence for a PCC in the ER membrane was obtained by electrophysiological experiments (Simon and Blobel, 1991). Shortly after, the channel function was assigned to the heterotrimeric Sec61 complex (Görlich and Rapoport, 1993). It consists of the central pore-forming Sec61 α subunit and two smaller auxiliary subunits Sec61 β and Sec61 γ . The structures of Sec61 and its bacterial and archaeal homologs (SecYEG and SecYE β respectively) have been studied extensively (see for example Van den Berg et al., 2004; Tsukazaki et al., 2008; Zimmer et al., 2008; Egea and Stroud, 2010; Gogala et al., 2014; Park et al., 2014; Voorhees et al., 2014; Voorhees and Hegde, 2016). The resulting structural models revealed a number of key structural elements which are universally conserved (see figure 1.1).

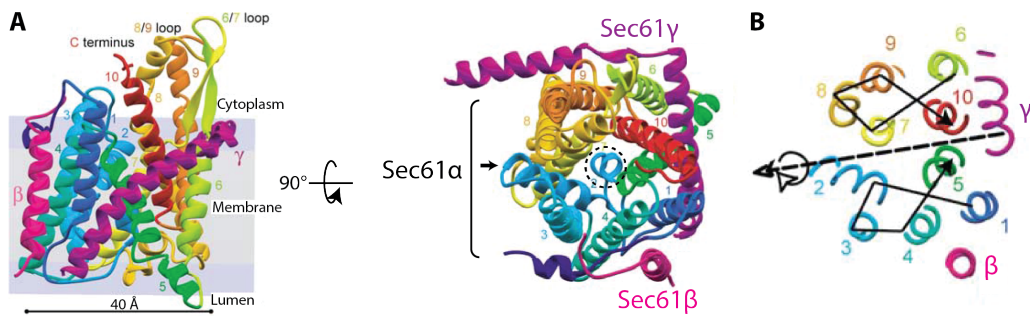


Figure 1.1 – Architecture of the heterotrimeric Sec61 channel. A) Model of the Sec61 channel, views from the membrane plane (left) and cytosol (right). In the cytoplasmic view, the lateral gate between TMHs 2b and 7 is marked by an arrow in the cytoplasmic view and the plug helix obstructing the central pore is indicated with a dashed circle. B) Sec61 α TMHs 1-5 and TMHs 6-10 form two pseudosymmetric halves. (Structural model from *Methanococcus janaschii*, figure adapted from Van den Berg et al., 2004).

Sec61 α (SecY)

The Sec61 α subunit is an integral membrane protein with ten α -helical transmembrane segments (TMS) forming an hourglass shaped cylinder with a central pore. Two connecting loops between transmembrane helix (TMH) 6 and 7 (L6/7) as well as TMH 8 and 9 (L8/9) extend into the cytoplasm. In its idle state, the Sec61 α central pore is blocked by a small helical segment (TMH 2a) therefore known as the plug domain. The pore ring - a circular arrangement of six hydrophobic residues at the constriction site - seals the channel against ion-leakage during the translocation process (Park and Rapoport, 2011). When viewed from the cytosol, the transmembrane (TM) portion of Sec61 α has a clam-shell like structure with two pseudosymmetric halves formed by TMH 1-5 and 6-10 respectively. The back side of the channel

where these two halves are connected by a luminal loop between TMH 5 and 6 is referred to as the hinge region. In the front the two halves meet at helices 2b and 7, an interface which has been coined as the channel's 'lateral gate'. This is because the N-terminal half of Sec61 α can undergo a hinge-like rigid body movement with respect to the C-terminus. This movement results in opening of the channel towards the lipid bilayer between TMHs 2b and 7. Both, SS as well as TMS of nascent peptide substrates have been found to be intercalated in the Sec61 lateral gate in corresponding cryo-EM and crystal structures (Gogala et al., 2014; Park et al., 2014; Voorhees and Hegde, 2016). Hence this site is currently seen as the canonical region for hydrophobic stretches to co-translationally partition into the membrane.

Sec61 γ (SecE)

The conserved core of Sec61 γ forms two α -helical segments. The N-terminal amphipathic helix resides perpendicular to the membrane plane on the cytoplasmic side and contacts the C-terminal half of the Sec61 α -subunit. It is connected to the C-terminal helix via a short hinge region. The C-terminal fragment spans the membrane in a diagonal manner traversing both halves of Sec61 α . Overall, this results in a clamp-like binding to the central channel.

Sec61 β (SecE)

Sec61 β is a tail-anchored protein and hence consists of a cytoplasmic N-terminal domain and a single C-terminal TMH. The cytoplasmic region is most likely disordered and has not been visualized in available structures. The TMH also appears to be rather dynamic but is in principle located in the vicinity of Sec61 α TMHs 1 and 4.

1.2.2 Biological functions of Sec61

Biochemical as well as structural studies have further provided important mechanistic insights into the co-translational action of Sec61. To date, four different functional states of Sec61 have been characterized: resting, primed, translocating and inserting.

The resting state most closely resembles the crystal structures of the isolated (archaea-) bacterial PCC. In this state the plug obstructs the pore and the lateral gate is completely closed (Van den Berg et al., 2004; Gogala et al., 2014). The main ribosome-Sec61 interaction is established between the backbone of the 28S rRNA,

the ribosomal protein eL29 and the uL23/uL29 universal adaptor site on the one hand and the conserved cytoplasmic L6/7 and L8/9 loops of Sec61 α on the other hand (Becker et al., 2009; Voorhees et al., 2014).

Upon ribosome-binding the channel adopts the primed state, indicated by slight opening of the lateral gate on the cytoplasmic side. The plug-domain, the pore-ring residues as well as the luminal portion of the channel remain unaltered (Voorhees et al., 2014).

When a nascent chain substrate engages the channel, hydrophilic stretches are translocated vertically across the ER membrane. This is possible by the plug becoming either disordered or being dynamically relocated and by a slight shift of TMHs 1 and 10 (Gogala et al., 2014; Voorhees et al., 2014).

When the PCC encounters a hydrophobic signal peptide (SP) or a TMS its lateral gate can open completely to allow direct passage of these segments into the lipid bilayer (inserting state, visualized by Voorhees and Hegde, 2016; Gogala et al., 2014). Two models for partitioning of hydrophobic segments into the membrane have been proposed. The most prevailing model is the 'in-out' model, where a helical segment first fully engages the central Sec61 cavity and then moves laterally into the lipid environment. In contrast to that stands the 'sliding model' proposed by the Von Heijne Lab (Cymer et al., 2015). Here, the hydrophobic segment contacts the lipid phase already early in the insertion process. This could possibly occur via the crack in the cytoplasmic face of the channels' lateral gate which is induced by ribosome binding. The segment then slides along the outer surface of Sec61 to be fully accommodated in the membrane. The extent of intercalation into the lateral gate might correlate to the segments' hydrophobicity with less hydrophobic segments being more likely to remain fully intercalated. The considerations behind the sliding model will be discussed in more detail in section 1.2.5.

The biological roles of the two small PCC subunits are less well characterized.

Sec61 γ (SecE) has mostly been studied in the bacterial system. It is essential for translocation and likely has stabilizing function for the core subunit. The functionally important portion lies in the Sec61 γ hinge and its C-terminal helix while the N-terminal amphipathic helix might solely be important for correct localization of the protein (Lycklama a Nijeholt et al., 2013; Kontinen et al., 1996; Murphy and Beckwith, 1994; Kihara et al., 1995).

In contrast, Sec61 β (SecG) is dispensable for translocation and insertion of PCC substrates (Kalies et al., 1998). Moreover there is apparently no sequence conservation

between Sec61 β and SecE proteins. The protein has been associated with the processes involved in handover of the ribosome-nascent-chain-complex (RNC) from SRP to the Sec61 channel. First, Sec61 β has been proposed to act as a guanine nucleotide exchange factor for the β -subunit of the SR (Helmers et al., 2003; Jiang et al., 2008). Second, delayed translocation upon Sec61 β deletion (Kalies et al., 1998) as well as crosslinking of nascent chains to Sec61 β during early translocation events (Laird and High, 1997; Meacock et al., 2002; MacKinnon et al., 2014; McKenna et al., 2016) imply a function in inserting the nascent chain into the tunnel pore. Besides, both Sec61 β and the prokaryotic SecE have been implicated to be essential for secretion of certain substrates (Kelkar and Dobberstein, 2009; Sibbald et al., 2010).

1.2.3 Accessory translocon factors

An array of different accessory factors are associated more or less transiently with the PCC *in vivo* to form a dynamic assembly, known as the translocon. They are implicated in different NC modifications or topology determination for nascent membrane proteins. The core translocon components are the translocating chain-associating membrane protein (TRAM), the translocon-associated protein complex (TRAP) and the oligosaccharyltransferase complex (OST).

The translocating chain-associating membrane protein (TRAM)

The integral membrane protein TRAM has been proposed to consist of eight (Tamborero et al., 2011) TMS. It is an essential translocon component for many secretory and transmembrane proteins (Görlich et al., 1992; Görlich and Rapoport, 1993). However, some substrates engage the Sec61 channel in the absence of TRAM. TRAM has been implicated in early translocation events in a signal-sequence dependent manner (Görlich et al., 1992; High et al., 1993; Mothes et al., 1994; Voigt et al., 1996). Furthermore, it has been proposed to have a functional role during membrane protein biogenesis, possibly chaperoning TMS with charged or hydrophilic residues (Görlich and Rapoport, 1993; Saurí et al., 2007; Heinrich et al., 2000; Meacock et al., 2002; McCormick et al., 2003; Devaraneni et al., 2011).

The translocon-associated protein complex (TRAP)

TRAP is a heterotetrameric complex in the ER membrane, formerly known as the signal sequence receptor complex (SSR) (Hartmann et al., 1993). The α - β - and δ -subunits have a single TMS whereas TRAP γ adopts a four-fold membrane spanning topology. The complex has been crosslinked to NCs upon translocation to the ER

lumen (Wiedmann et al., 1987; Görlich et al., 1990). More recently it has also been indicated to be involved in topology determination of membrane proteins (Sommer et al., 2013). Moreover, mutations in the TRAP δ -subunit result in defects in asparagine-linked glycosylation due to skipping of some glycosylation sites in the nascent chain (Losfeld et al., 2014). Low resolution cryo-EM structures as well as cryo-electron tomography (CET) followed by subtomogram averaging have identified TRAP as a stoichiometric component of mammalian translocon complexes and have revealed its positioning with respect to Sec61 and the ribosome (Ménétret et al., 2005; Ménétret et al., 2008; Pfeffer et al., 2014; Pfeffer et al., 2015; Pfeffer et al., 2017). The TRAP transmembrane region is flanking the C-terminal half of the PCC. The complex makes contact to the large ribosomal subunit via its cytoplasmic domain. On the other side of the membrane the TRAP complex harbors a large luminal lobe, which resides below the Sec61 central pore and in close proximity to the Sec61 hinge region and the luminal segment of the OST. The complex has been unamenable to high resolution analysis by single-particle cryo-EM, likely because of its tendency to be disordered or partly dissociate upon detergent solubilization.

The oligosaccharyltransferase complex (OST)

The eukaryotic oligosaccharyltransferase complex (OST) is a multisubunit complex and an integral part of the ER membrane. It is responsible for the most abundant post-translational modification of nascent secretory proteins: asparagine-linked glycosylation (N-glycosylation). OST scans the emerging polypeptide for glycosylation sequons (Asn-X-Ser/Thr(/Cys), where X is any aa except proline, Gavel and Heijne, 1990) and covalently links a preassembled oligosaccharide to the side chain nitrogen of the Asn residue by an N-glycosidic bond (Helenius and Aebi, 2004; Mohorko et al., 2011; Shrimall et al., 2015). The enzyme's catalytic center is located in the ER lumen (Hanover and Lennarz, 1980; Welply et al., 1983). N-glycosylation has been shown to be an important determinant for protein topology, since it prevents retrotranslocation of the modified peptide stretch, thereby constraining it to the luminal compartment (Welply et al., 1983). Furthermore, naturally occurring as well as genetically engineered glycosylation motifs can be used diagnostically, to probe the location of a certain protein region (Nilsson and Von Heijne, 1993). This is possible because successful modification results in a detectable shift of molecular weight (MW) of approximately 2-3 kilodalton (kDa). The shift can be reversed by treating the sample with endoglycosidase H (Endo H) to enzymatically remove the oligosaccharide.

1.2.4 Membrane protein terminology

Different concepts are used to classify membrane proteins which are inserted into the bilayer in a Sec61-dependent manner. First, they can be separated into bitopic versus polytopic (also: single-spanning vs. multi-spanning) proteins (Blobel, 1980). A bitopic protein has only a single TMH in its mature form whereas several TMH traverse the membrane multiple times in polytopic substrates.

Another classification approach clusters TMPs into type I, II or III, based on the final orientation of the most N-terminal TMH as well as the presence or absence of a cleavable N-terminal signal peptide (Figure 1.2) (Von Heijne and Gavel, 1988; Spiess, 1995).

Type I TMPs harbor such a signal and the subsequent TMH anchors the mature pro-

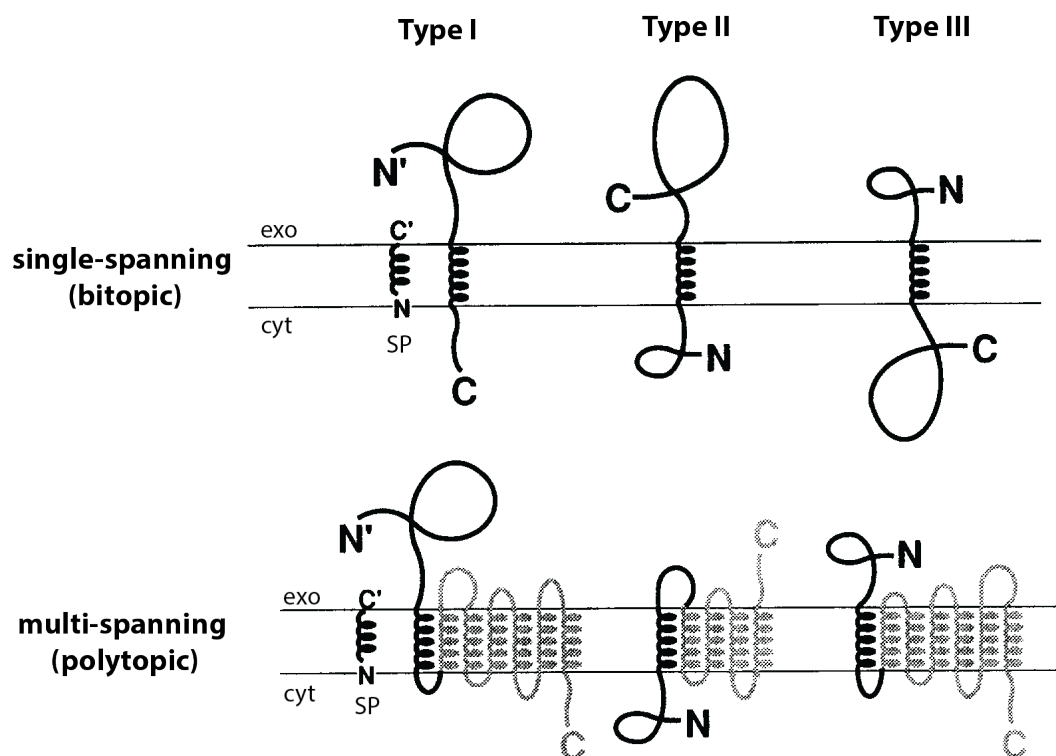


Figure 1.2 – Types of membrane proteins. Membrane proteins are be classified according to their number of TMS (single- vs. multispanning) or based on the orientation of the most N-terminal TMS in the mature protein in combination with the presence or absence of a cleavable N-terminal signal peptide (SP) (Adapted from Spiess, 1995).

tein in the membrane with its N-terminus in the lumen and its C-terminus in the cytosol (N_{lum} - C_{cyt}) once the signal sequence is cleaved off.

Type II TMPs lack a cleavable SS and are instead targeted via their most N-terminal TMH which then serves as a signal-anchor (SA) in the ER membrane. In type II proteins, the N-terminus resides in the cytoplasm while the SA-C-terminal-region is

translocated into the lumen ($N_{cyt}-C_{lum}$).

Some type II TMP initially expose their N-terminus to the lumen (head-first insertion) and subsequently reorient the N-terminal TMH in the membrane plane to obtain their mature topology (Devaraneni et al., 2011). Others are more likely to retain their N-terminus on the cytoplasmic site, resulting in a looped (hairpin) conformation during the insertion process (MacKinnon et al., 2014). Both, hairpin and head-first insertion followed by reorientation also seem to be possible trajectories for cleavable signals of type I TMP to reach their looped pre-cleavage orientation (Shaw et al., 1988; Rösch et al., 2000; Rutkowski et al., 2001; Goder and Spiess, 2003; Vermeire et al., 2014). Hairpin vs. head-first insertion has been proposed to be a function of the number of aminoacids preceding the targeting signal at the N-terminus, with longer N-terminal stretches promoting the looped insertion (Kocik et al., 2012; MacKinnon et al., 2014; McKenna et al., 2017).

Finally, type III TMP also do not contain a cleavable signal. In contrast to type II proteins, they always insert in a head-first manner and do not reorient in the membrane. Thus, their N-terminus is being translocated across the membrane and the C-terminus of the SA points towards the cytosol ($N_{lum}-C_{cyt}$).

1.2.5 Sec61-dependent membrane protein insertion - a thermodynamic view

The sliding model for TMS insertion described in 1.2.2 was derived by evaluating the insertion process from a thermodynamic perspective (Cymer et al., 2015). Thermodynamically, a reaction is favorable, when it is connected with a negative difference in free energy (ΔG). According to the Gibbs equation (Equation 1.1), contributions to ΔG include changes in enthalpy ΔH as well as entropy ΔS for a certain temperature T .

$$\Delta G = \Delta H - T\Delta S \quad (1.1)$$

For insertion of polypeptides into the lipid bilayer, ΔG is influenced by partitioning of the peptide backbone as well as the aa side chains. In an extended conformation, backbone contributions are very unfavorable due to dehydration of hydrophilic peptide bonds. These ΔG penalties can be tremendously reduced, albeit not entirely abolished by intramolecular saturation of hydrogen bonds through adapting an α -helical conformation (Ladokhin and White, 1999; Almeida et al., 2012). Partitioning of hydrophobic side-chains on the other hand is beneficial for ΔG of peptide insertion. In a simplified model, a polypeptide segment is preferentially inserted

into the bilayer if the hydrophobic effect from the side-chains can outbalance the unfavorable contributions from insertion of the α -helical peptide backbone. Hence, the exact ΔG for lipid-bilayer insertion depends on the aa sequence of a given peptide stretch. A biological hydrophobicity scale has been derived to rank amino acids according to their contributions to ΔG depending on their position in a peptide segment (Hessa et al., 2005). Factors which influence a residue's contribution include its hydrophobicity, charge as well as helix-forming propensity.

Assuming a negative ΔG , it should in principle be possible for a TMS to insert into the bilayer spontaneously, without the help of the PCC. Therefore, the primary role of the channel in the insertion might be to facilitate initial contact with the hydrophobic membrane core. The lateral gate could then open gradually as the TMS slides along the PCC surface and provide a protected environment for less hydrophobic faces of TMS or α -helical signal sequences. In addition, the PCC as well as other translocon factors could impact the topology and tertiary structures of membrane proteins.

1.2.6 TMH topology determinants

Several determinants have been identified to influence the topology of transmembrane helices (For review see for instance Higgy et al., 2004; Lee and Kim, 2014). The first important factor is the charge distribution along the THM. Usually the TMH terminus with the more positive net charge preferentially locates to the cytoplasmic face of the membrane ('positive inside rule', Von Heijne, 1986). Charges of immediate flanking residues also contribute to this rule. Another topogenic dimension is given by the segment's hydrophobicity. This includes overall hydrophobicity, length of the hydrophobic region as well as the distribution of hydrophobic residues along the TMS. More hydrophobic segments seem to integrate more readily whereas marginally hydrophobic segments have been proposed to have longer retention times at the PCC. In some cases hydrophobicity is in direct competition with the positive inside rule. It has for example been shown that increased hydrophobicity prevented the reorientation of a type II SA after headfirst insertion, despite a net positive N-terminal charge (Goder and Spiess, 2003). Furthermore, folding of N-terminal domains will prevent a downstream segment from adopting a N_{lum} - C_{cyt} topology. The reason for this is that the PCC is only capable of translocating extended polypeptides but not stably structured regions across the membrane.

For polytopic transmembrane proteins, the situation is even more complex and by far less well understood. Initially, it was believed that the orientation of the most

N-terminal TMH determines the topology of the entire protein (Blobel, 1980). However more recent data shows that topogenic signals can be present in several regions of the nascent chain and integration of TMHs can be interdependent or strongly influenced by adjacent sequences. Recently, up to 100 residues C-terminal of a TMS have been indicated to influence the topology of this segment (Junne and Spiess, 2017). Similarly, it is still poorly understood at what point a helix disengages from the PCC. TMHs can be released one-by-one in a series of alternating translocation and membrane integration steps (linear insertion model, Blobel, 1980; Mothes et al., 1997). Alternatively, some TMS have been proposed to be retained in the vicinity of Sec61 and interact with more C-terminal TMS prior to final release (Borel and Simon, 1996; Meacock et al., 2002; Ismail et al., 2006; Cross and High, 2009). Finally, certain TMH might even transiently disengage the PCC, only to return to the lateral gate region later during their biogenesis to aid insertion of more C-terminal segments (Heinrich and Rapoport, 2003).

An example for a class of TMHs which often fail to successfully and stably integrate into the lipid bilayer themselves are marginally hydrophobic segments. Instead, they can depend on properties of flanking loops and insertion of neighboring helices (Hedin et al., 2010; De Marothy and Elofsson, 2015). In accordance with this, helices can be repositioned during the insertion process (Lu et al., 2000; Kanki et al., 2002; Kauko et al., 2010; Watson et al., 2013).

Another feature of more complex transmembrane proteins are segments which adopt non-standard topologies in the mature protein. Examples for these are extra-long TMS which span the membrane diagonally, re-entrant or very short TMS which only traverse the membrane incompletely, or amphipathic helices, which lie perpendicularly on one face of the membrane (Von Heijne, 2006; De Marothy and Elofsson, 2015; Tsirigos et al., 2018).

Moreover, some proteins exist in multiple topogenic forms and it has even been shown for a bacterial polytopic substrate that topology can be reversed completely upon exchange of a single residue (Seppälä et al., 2010).

The exact role of the Sec61 complex and other translocon factors during the topogenesis of TMPs remains unclear. Also, the topogenic signals within polytopic TM proteins and the translocon mediated responses remain only vaguely characterized.

1.3 Biogenesis of bovine opsin

Bovine opsin is a prototypical G-protein coupled receptor (GPCR). This class of integral membrane proteins adopts a multispanning topology with seven TMS. Opsin is highly conserved across organisms and serves as a photon receptor in the retina after binding to its essential co-factor 11-*cis*-retinal.

The crystal structure of bovine rhodopsin (i.e. opsin in covalent linkage to the chromophore) provided detailed insights into the protein's conformation and topology (see figure 1.3) (Palczewski et al., 2000; Palczewski, 2006).

Opsin is a polytopic type III membrane protein and hence adopts a N_{lum} - C_{cyt} topol-

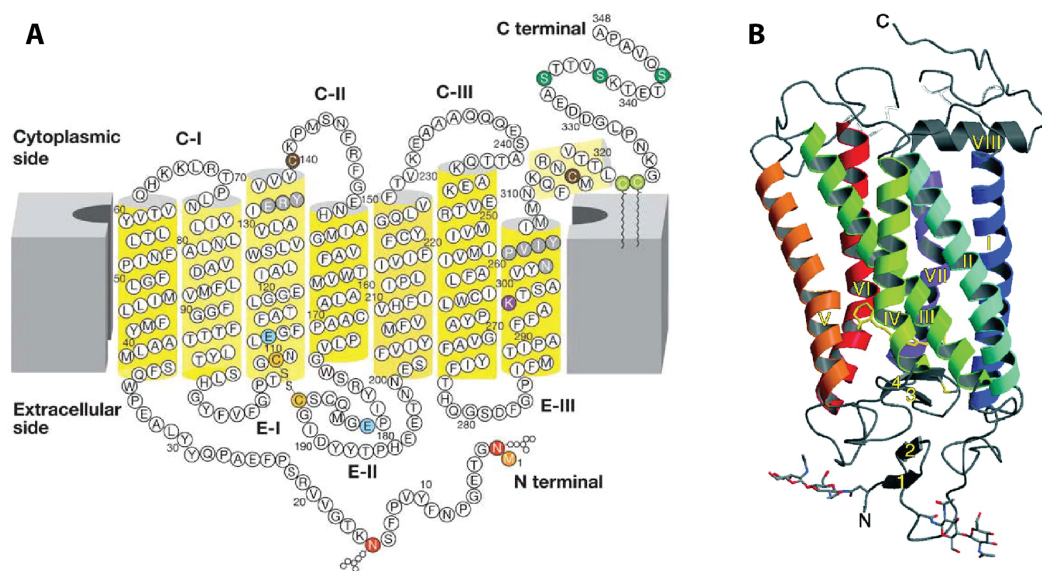


Figure 1.3 – Topology and structure of bovine rhodopsin. A) Schematic depiction of bovine rhodopsin topology. Functionally important residues (grey, blue, purple) as well as residues carrying post-translational modifications (orange, red, yellow, light and dark green) are highlighted. B) Crystal structure of bovine rhodopsin. (Adapted from Palczewski et al., 2000; Palczewski, 2006).

ogy during insertion with the first TMS serving as a SA. This topology has been confirmed in its native location in the rod cell disk membrane and when translated *in vitro* in the presence of canine microsomal ER membranes (Goldman and Blobel, 1981; Hargrave et al., 1983). Its N-terminal tail harbors two N-glycosylation sites at N2 and N15 (Hargrave, 1977).

Opsin has a long standing tradition as a model substrate for studying both co-translational membrane protein insertion as well as N-glycosylation (Hargrave, 1977; Goldman and Blobel, 1981). It has been demonstrated, that opsin TMSs 1-3 as well as 5-7 constitute stable individually folded subdomains (Ridge et al., 1995;

Ridge et al., 1996). In line with that, crosslinking experiments suggest that these groups of TMH form independent Sec61-adjacent bundles prior to release into the bilayer (Ismail et al., 2006; Ismail et al., 2008). In contrast, opsin TMH 4 rapidly leaves the PCC independent of sequence context. These and earlier studies use *in vitro* translated cys-null variants of opsin with a single cysteine residue at a central position within one of the seven TMS. Crosslinking of this cysteine to adjacent proteins is used to explore the interactions of the crosslinked segment with different translocon components upon varying nascent chain length (Laird and High, 1997; Meacock et al., 2002; Ismail et al., 2006; Ismail et al., 2008).

Based on crosslinks to Sec61 α and Sec61 β a two-phase interaction model has been suggested. In phase I, when a TMH first encounters the PCC, it can be simultaneously crosslinked to both Sec61 subunits. With longer chain lengths, crosslinks to the same position in the opsin chain can only be detected with Sec61 α indicating that the corresponding TMS has now moved to a different environment in the translocon region. Furthermore, the authors discovered a protein associated with the ER translocon of 10 kDa, therefore named PAT-10, which interacts stably with opsin TMH 1 and 5 throughout the biogenesis of the protein (Meacock et al., 2002; Ismail et al., 2006; Ismail et al., 2008). Also, they show that TMH 2 undergoes transient interactions with the TRAM protein (Meacock et al., 2002). Based on these data, the authors have developed a model for the translocon interactions during opsin biogenesis (see Figure 1.4).

The crosslinking experiments serve as a solid basis for understanding the processes involved in the insertion process of the polytopic opsin substrate. However, they can only provide very limited spatial information on the position of individual TMS during the biogenesis. Besides, the authors did not investigate the role of OST which performs N-glycosylation on two sites in the opsin N-terminal region at the same time as the protein is being inserted co-translationally.

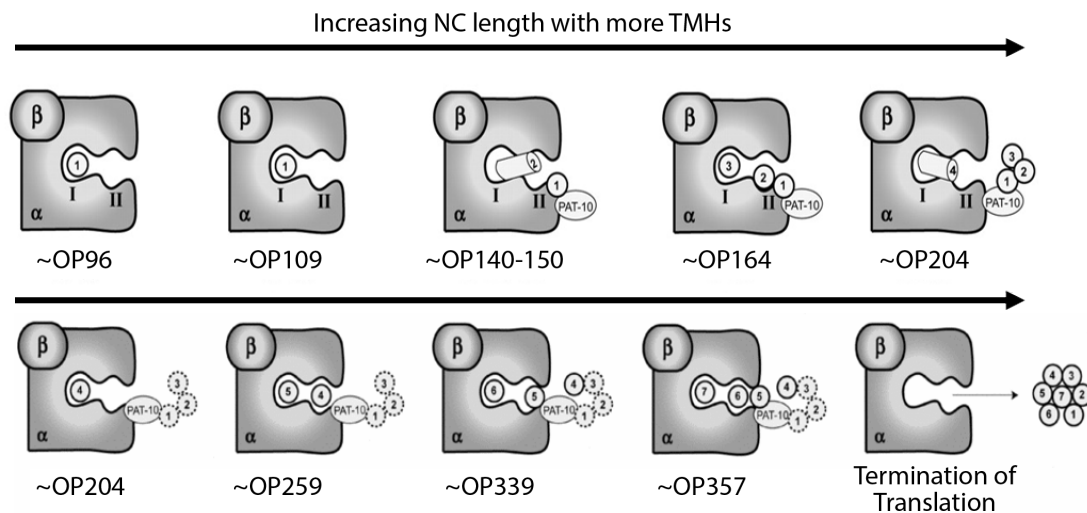


Figure 1.4 – Bovine opsin biogenesis. Schematic depiction of opsin biogenesis including interactions with specific translocon components (Adapted from Ismail et al., 2006; Ismail et al., 2008).

1.4 Asparagine-linked glycosylation at the ER membrane

1.4.1 Evolution and significance of OST

The vast majority of proteins which enter the secretory pathway are N-glycosylated by the OST enzyme. Homologs of the catalytic subunit can be found in all three kingdoms of life (STT3 proteins in eukaryotes, see Kelleher et al., 2003; AglB in archaea, see Spirig et al., 1997; PglB in bacteria, see Szymanski et al., 1999). The bacterial species can be found in the inner leaflet of the plasma membrane, while the eukaryotic enzyme is located in the ER membrane. An evolutionary perspective reveals that the number of N-glycosylation-sites in an organism's proteome has increased drastically from bacteria over fungi to humans (Dell et al., 2010; Zielinska et al., 2012). Likewise, the complexity of the modifying enzyme has increased dramatically. In bacteria, OST is a single subunit enzyme. In *Saccharomyces cerevisiae* it is a heterooctameric complex consisting of the following subunits (homologs in higher eukaryotes with corresponding abbreviations in parentheses) (Karaoglu et al., 1997; Spirig et al., 1997; Cherepanova et al., 2016):

- Oligosaccharyltransferase protein 1, Ost1p (ribophorin I, RPN1)
- Oligosaccharyltransferase protein 2, Ost2p (defender against apoptotic cell-death 1, DAD1)

- Oligosaccharyltransferase protein 3 or 6, Ost3p or Ost6p (magnesium transporter 1/implantation-associated protein, MagT1/IAP; Tumor suppressor candidate 3, TUSC3/N33; DC2)
- Oligosaccharyltransferase protein 4, Ost4p (OST4)
- Oligosaccharyltransferase protein 5, Ost5p (transmembrane protein 258, TMEM258)
- Wheat germ binding protein 1, Wbp1p (Ost subunit of 48 kilodalton, OST48)
- Suppressor of Wbp1 protein, Swp1p (ribophorin 2, RPN2)
- Stt3p (staurosporin and temperature sensitive mutant 3, STT3A and STT3B).

Five of the subunits are encoded by essential genes in *S. cerevisiae* (OST1, OST2, WBP1, SWP1 and STT3).

By gene duplication metazoans have evolved a second paralog of the catalytic subunit. The resulting STT3A and STT3B proteins assemble with a partially overlapping set of auxiliary factors (RPN1, RPN2, OST4, TMEM258, DAD1, OST48). In addition, the STT3A- and B complexes incorporate paralog specific subunits (DC2 and KCP2 [keratinocyte associated protein 2] for STT3A; MagT1/IAP or Tusc3/N33 for STT3B) (Figure 1.5).

N-glycosylation has been shown to play a pivotal role in glycoprotein quality con-

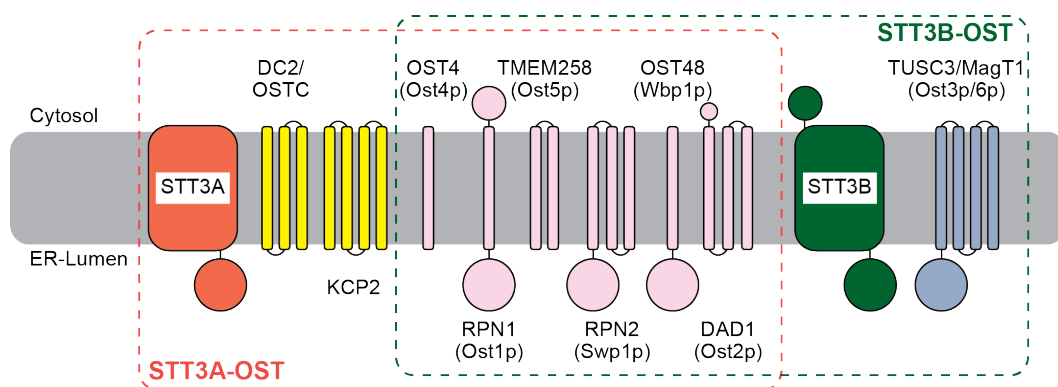


Figure 1.5 – Composition of eukaryotic OST complexes. Eukaryotic OST consists of eight to nine different subunits. Higher eukaryotes express two complex isoforms which assemble around paralogous versions of the catalytic STT3A/STT3B subunits (names of *S. cerevisiae* homologs in parentheses) (Adapted from Braunger et al., 2018).

trol, protein folding as well as -trafficking (Sato et al., 2012; Tannous et al., 2015; Caramelo and Parodi, 2015). Concomitantly, defects in N-glycosylation in humans

result in a class of diseases called congenital disorders of glycosylation (Reviewed for example in Hennet and Cabalzar, 2015).

1.4.2 The oligosaccharide precursor

The oligosaccharide precursor is assembled in a series of glycosyltransferase reactions. They are catalyzed by alg-family proteins (alg = asparagine-linked glycosylation). Via a pyrophosphate (PP_i) bond the precursor is linked to a dolichol molecule (Dol), which anchors it to the ER membrane. The first seven sugar residues are assembled on the cytoplasmic face of the rough ER. The partially assembled precursor is then flipped across the membrane by Rft1 (Helenius et al., 2002) and its synthesis is completed in the ER lumen. The final lipid-linked oligosaccharide (LLO) adopts a branched structure and consists of two N-acetyl-glucosamines, nine mannose units and three glucose moieties ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-PP-Dol}$) (Figure 1.6).

Once the preassembled oligosaccharide has been transferred to the substrate pep-

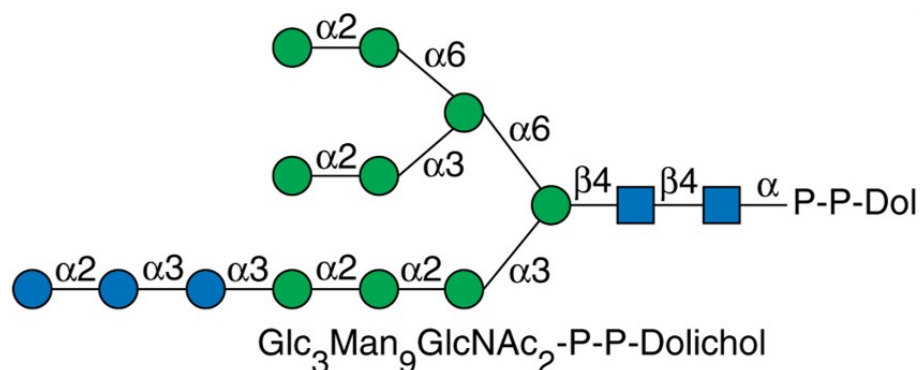


Figure 1.6 – The LLO precursor. Schematic depiction of the lipid-linked oligosaccharide precursor (LLO) including the types of glycosidic linkage between monosaccharide components (Adapted from Stanley et al., 2015).

tide, it undergoes several trimming and modification steps. These lead to a great variety of N-linked glycans which play crucial functional and structural roles.

1.4.3 Catalytic mechanism of STT3-type enzymes

The crystal structures of bacterial PglB (protein glycosylation B) and archaeal AglB (archaeal glycosylation B) single-subunit OST enzymes have been solved in different functional states (Igura et al., 2008; Maita et al., 2010; Lizak et al., 2011; Matsumoto et al., 2013; Matsumoto et al., 2017; Napiórkowska et al., 2017). A high degree of structural conservation has been observed despite low levels of sequence conservation,

in particular for the transmembrane region and the functionally important core of the C-terminal domain. Moreover, the acceptor sequon requirements (N-X-S/T(/C), see Gavel and Heijne, 1990; extended to E/D-X-N-X-S/T in bacteria, see Kowarik et al., 2006) as well as the chemical nature of the LLO (an isoprenoid membrane anchor and a pyrophosphate leaving group coupled to the preassembled glycan) are conserved widely across all kingdoms of life. This suggests that mechanistic implications from the PglB and AglB structures also apply to the catalytic STT3-type subunits of eukaryotic OST.

Overall, the structures are composed of a large N-terminal transmembrane domain and a slightly smaller C-terminal soluble domain which could also be crystallized separately (Igura et al., 2008; Maita et al., 2010). Furthermore, the OST crystal structures reveal two groves at the membrane interface on opposite sides of the enzyme. They serve as binding regions for the two types of substrates which need to be handled by OST in a coordinated manner: the acceptor peptide and the LLO. The two cavities are connected by a tunnel (also referred to as the enzyme's "loophole") to allow for covalent substrate linkage. The acceptor sequon must be localized in a flexible loop in order to accommodate in the active site, thus precluding modification of folded peptide regions.

Hence, the structures defined three spatially separated functional sites for binding of the peptide sequon, accommodation of the LLO donor as well as an active site region. Individual residues at each of these sites which have been shown to be important for OST functionality are listed in table 1.1 for different organisms.

In more detail, the N-terminus of the enzyme adopts a unique fold of thirteen transmembrane segments. Most connecting cytoplasmic and external loops (EL) are relatively short. Exceptions to this are EL1 (between TMS 1 and 2) and EL5 (between TMS 9 and 10). EL1 is well structured in all available models and contains two α -helical stretches.

In contrast, EL5 adopts different conformational states depending on absence or presence of LLO and peptide substrates in the crystal structures. The loop is completely structured in the enzyme's apo state (absence of peptide and LLO, pdb 3wak) (Matsumoto et al., 2013) or when both, an acceptor peptide and a synthetic, nonhydrolyzable LLO analog are tightly bound (pdb 5ogl) (Napiórkowska et al., 2017). In these structural models, it forms a short helical segment in the N-terminal half (N-EL5). In contrast, the C-terminal region does not contain defined secondary structural elements despite being rigid enough to be visualized in the crystal structures. When bound to a substrate peptide but not the glycan donor (pdb 3rce, 5gmy)

Table 1.1 – Functionally important residues in STT3-type enzymes of different organisms. Residues which have been shown to be involved in crucial interactions for substrate binding or catalysis in crystal structures of single subunit OST enzymes (PglB, AglB) (Igura et al., 2008; Maita et al., 2010; Lizak et al., 2011; Matsumoto et al., 2013; Matsumoto et al., 2017; Napiórkowska et al., 2017) and corresponding residues in eukaryotic STT3-type enzymes (Stt3p, STT3A, STT3B). Conserved motifs are underlined. Ser/Thr = serine/threonine residues in +2 position of the acceptor sequon, PP_i = pyrophosphate, NAc = N-acetyl group of the reducing-end GlcNAc moiety, M²⁺ = divalent metal ion (Mn²⁺ *in vivo*), Asn = acceptor asparagine.

Function	Binding partner	<i>C. lari</i> (PglB)	<i>A. fulgidus</i> (AglB)	<i>S. cerevisiae</i> (Stt3p)	<i>H. sapiens</i> (STT3A)*	<i>H. sapiens</i> (STT3B)*
Peptide binding	<u>WWD</u>					
	Ser/Thr	W463	W550	W516	W525	W604
	Ser/Thr	W464	W551	W517	W526	W605
	Ser/Thr	D465	D552	D518	D527	D606
	<u>DK/MI</u>					
	Ser/Thr	I572	K618	K586	K595	K674
LLO binding	PP _i	Y196	W215	W208	W209	W263
	PP _i	R375	R426	R404	R405	R459
	NAc	Y468	H555	Y521	Y530	Y609
Active site	M ²⁺ / Asn	D56	D47	D47	D49	D103
	M ²⁺	D154	D161	D166	D167	D221
	M ²⁺	D156	H163	E168	E169	E351
	<u>TIXE/SVSE</u>					
	M ²⁺ / Asn	E319	E360	E350	E351	E405

*residue numbers are identical to *Canis lupus familiaris* STT3A/B which is used in the present study

(Lizak et al., 2011; Matsumoto et al., 2017), N-EL5 appears flexible and cannot be visualized. Finally, the loop is completely disordered in an archaeal structure which did not contain any substrates or analogs but showed extra density for a sulfate-ion which was proposed to mimick the pyrophosphate group of the LLO precursor (pdb 3waj) (Matsumoto et al., 2013). Besides, tethering of N-EL5 to the transmembrane region by crosslinking to the C-terminal, luminal end of TMH 11 has significantly reduced turnover of a fully assembled oligosaccharide precursor while leaving transfer of a minimal monosaccharide substrate unaffected. Based on these finding it was proposed that N-EL5 immobilization impairs LLO binding (Napiórkowska et

al., 2017). Hence, EL5 has been assigned a prominent role in the catalytic mechanism.

TMS 1-4 as well as 10-13 form the most prominent contact sites to the well-structured C-terminal periplasmic domain.

The periplasmic domain contains the conserved WWD and DK/MI motifs which form the binding pocket for the +2 threonine/serine residue in the acceptor peptide via hydrogen bonds (WWD motif) or van der Waals interactions (DK/MI motif) (Igura et al., 2008; Maita et al., 2010; Lizak et al., 2011). This ensures spatial separation from the modified asparagine which can in turn accommodate in the loop-hole. The catalytic pocket is also highly conserved. Four acidic residues (D56, D154, D156, E319 in *C. lari*, for other organisms see table 1.1), coordinate a divalent metal ion (M²⁺) which is required for catalysis and has been proposed to have a post-catalysis stabilizing role for the lipid-pyrophosphate (Lizak et al., 2011; Napiórkowska et al., 2017). D56 and E319 also interact with the acceptor asparagine, resulting in activation of its nitrogen and facilitating nucleophilic attack on the LLO. E319 is located in the C-terminal half of EL5 and is part of the widely conserved TIXE (Thr-Ile-X-Glu) motif in archaea and eubacteria, or the corresponding SVSE (Ser-Val-Ser-Glu) motif among eukaryotes (Matsumoto et al., 2013).

Based on the structures in different functional states the following catalytic cycle has been proposed (Matsumoto et al., 2013; Napiórkowska et al., 2017):

- Step 1: apo-state
- Step 2: polypeptide-bound- or LLO-bound-state
- Step 3: ternary complex OST-polypeptide-LLO
- Step 4: transition state, catalysis, glycopeptide release
- Step 5: lipid-pyrophosphate-bound state

Napiórkowska et al. have set up the hypothesis that LLO binding is likely to occur prior to peptide binding (Napiórkowska et al., 2017). The reasoning is that EL5-C engagement upon peptide binding might limit the access of the bulky oligosaccharide precursor to its binding site.

1.5 The eukaryotic OST complex

1.5.1 Complex architecture and distance restraints

Structural data on eukaryotic OST and its subunits remains scarce. Therefore, domain assignments and assumed topologies rely almost exclusively on structure predictions and biochemical characterization, for example by protease protection studies or glycosylation mapping assays. Also, the arrangement of OST subunits with respect to each other remains elusive.

However, biochemical studies on the yeast OST complex have indicated existence of three OST subcomplexes (Karaoglu et al., 1997). Subcomplex I is formed by Ost1p and Ost5p. Subcomplex II contains the catalytic Stt3p subunit, Ost4p as well as Ost3p or Ost6p respectively. Subcomplex III encompasses Swp1p, Wbp1p and Ost2p. In metazoans these subcomplexes would correspond to RPN1 and TMEM258 (Subcomplex I); STT3A/B, OST4 and the paralog specific subunits (Subcomplex II); RPN2, OST48 and DAD1 (Subcomplex III).

In higher eukaryotes, distinct functions of the OST complexes assembling around STT3A or STT3B (from now on referred to as STT3A-OST and STT3B-OST respectively) have also been inferred from biochemical studies. These data have indicated that STT3A-OST but not STT3B-OST associates with the ribosome-Sec61-complex (Shibatani et al., 2005; Ruiz-Canada et al., 2009). The overall spatial arrangement of the mammalian OST-containing ribosome-translocon-complex (RTC) has been established by moderate resolution cryo-electron tomography structures (see figure 1.7) (Pfeffer et al., 2014; Pfeffer et al., 2015; Pfeffer et al., 2017). They revealed that the OST TM region is localized proximal to the N-terminal half of Sec61 and the PCC's hinge region. In addition, OST has a large luminal extension, protruding approximately 90 Å into the ER lumen (Pfeffer et al., 2014). Ribosome associated nascent chains first have access to the OST active site when the distance between the PTC and the acceptor asparagine is 65-75 aa (Whitley et al., 1996). With respect to the membrane, the minimal distance for successful glycosylation between a signal sequence or TMD and the acceptor arginine has been shown to be 10-14 aa, corresponding to 20-30 Å, depending on the peptide conformation (Nilsson and Von Heijne, 1993; Nilsson et al., 1994).

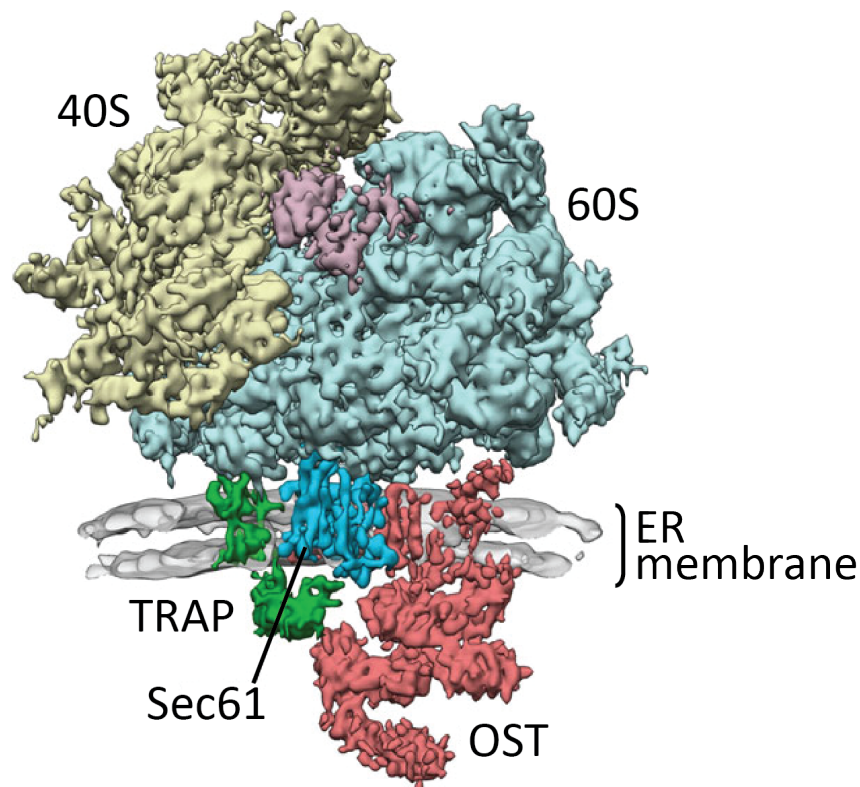


Figure 1.7 – CET structure of mammalian OST-containing RTCs. Subtomogram average of the mammalian OST-containing RTC filtered to 9.0 Å resolution reveals the spatial arrangement of translocon components (Adapted from Pfeffer et al., 2015).

1.5.2 The catalytic STT3 subunit

For a long time the number of TMH in eukaryotic STT3-type enzymes was not unambiguously known, ranging from 11 to 13 according to hydropathy plots and glycosylation mapping assays (Kim et al., 2005). However, recent evidence could confirm thirteen membrane spanning stretches in the eukaryotic enzyme (Lara et al., 2017) in analogy to the archaeal and bacterial homologs. Despite their shared topology and very high degree of sequence conservation (60% sequence identity), STT3A and STT3B display certain differences in their cellular roles (Kelleher et al., 2003; Ruiz-Canada et al., 2009). Functional analyses of the complex isoforms rely on selective purification of either STT3A- or STT3B-OST (Kelleher et al., 2003), siRNA mediated knock-down (Wilson and High, 2007; Ruiz-Canada et al., 2009) or CRISPR-mediated knock-out of either one of the catalytic STT3 paralogs (Cherepanova and Gilmore, 2016).

STT3A

The 81 kDa STT3A paralog forms the catalytic core of the STT3A-OST complex which acts purely co-translationally and is responsible for the majority of all N-glycosylation events (Ruiz-Canada et al., 2009). It has very stringent substrate requirements for both, the acceptor peptide as well as the LLO donor. This leads to skipping of glycosylation consensus sites under certain circumstances. First, sequons in close proximity to a signal sequence cleavage site have been observed to be unamenable to glycosylation by STT3A-OST (Ruiz-Canada et al., 2009). Sequon skipping can also be caused by closely spaced NXS acceptor sites (Shrimal and Gilmore, 2013). Besides, glycosylation by STT3A is very limited when the corresponding motif is found in a very C-terminal region of the substrate polypeptide (Shrimal et al., 2013). Finally, sub-optimal sequons, including those with internal cysteine residues are omitted at a high frequency by STT3A-OST (Malaby and Kobertz, 2014; Cherepanova et al., 2014).

STT3B

The OST complex containing STT3B can act co- and post-translationally (Kelleher et al., 2003; Ruiz-Canada et al., 2009). The major role proposed for STT3B-OST is glycosylation of substrates which have been skipped by STT3A-OST to ensure maximum coverage (Ruiz-Canada et al., 2009). In line with that, STT3B-OST is more active than STT3A-OST and more promiscuous in substrate and donor selection (Kelleher et al., 2003). STT3B has been shown to be more similar to yeast Stt3p than the STT3A paralog (Shrimal and Gilmore, 2013; Zielinska et al., 2010; Zielinska et al., 2012). In contrast to the yeast enzyme and the STT3A paralog, the 94 kDa STT3B features additional N- and C-terminal extensions as well as a significantly longer loop between TMH 12 and 13 (L12/13) (Kelleher et al., 2003).

Interestingly, STT3B-mediated hyperglycosylation of misfolded protein substrates at consensus sites which are neglected in the correctly folded proteins can target these substrates for ER-associated degradation (Sato et al., 2012; Cherepanova et al., 2016).

1.5.3 Shared accessory subunits

The majority of metazoan OST subunits form part of both complex isoforms.

RPN1 (Ost1p)

Ribophorins were originally named based on their localization to ribosome-binding

sites in the rough ER (Kreibich et al., 1978). RPN1 (68 kDa) was among the first three identified OST subunits, together with RPN2 and OST48 (Kelleher et al., 1992). RPN1 has been predicted to consist of a large luminal domain, a single transmembrane helix and a C-terminal cytoplasmic domain of approximately 150 aa (Crimaudo et al., 1987; Mohorko et al., 2011). Antibodies against the cytoplasmic domain inhibit protein translocation by preventing ribosome targeting to the membrane (Yu et al., 1990). RPN1 is crucial for glycosylation in a substrate dependent manner. Certain secretory and polytopic substrates can be successfully glycosylated in absence of RPN1 whereas the subunit plays an important role for modification of single-spanning substrates (Wilson et al., 2005; Wilson and High, 2007; Wilson et al., 2008).

RPN2 (Swp1p)

Primary structure analyses indicate that the RPN2 protein has a large luminal domain followed by three transmembrane helices at the C-terminus (Crimaudo et al., 1987). Notably, mammalian RPN2 displays a large difference in molecular weight compared to the homologous yeast protein (67 kDa RPN2, 32 kDa Swp1p) which is not true for any of the other OST subunits (Shibatani et al., 2005).

OST48 (Wbp1p)

OST48 is a type I single-spanning TMP with a substantial N-terminal luminal portion and only very few cytoplasmic residues downstream of the TMH. The protein has been suggested to act as a bridge between the ribophorin subunits via its luminal domain (Fu et al., 1997). Besides, the yeast homolog Wbp1 (Te Heesen et al., 1992; Silberstein et al., 1992) has been implicated in LLO precursor binding (Pathak et al., 1995). This suggestion is in line with the finding of a homology search which has identified a GIFT domain (for GldG/IFT = Gliding motility protein G/intraflagellar transport) in the Wbp1p and OST48 luminal domains which was proposed to have sugar binding function (Beatson and Ponting, 2004). However, confirmatory data for the LLO-binding hypothesis is lacking (Mohorko et al., 2011).

DAD1 (Ost2p)

Together with the catalytic subunits and the initially identified RPN1, RPN2 and OST48 subunits, DAD1 - a small 12 kDa protein - constitutes the fifth essential component of eukaryotic OST (Kelleher and Gilmore, 1997; Fu et al., 1997). It is also the OST subunit with the highest degree of sequence conservation between mammalian species (100% identity between human, mice, rat, and hamster; 99% among human,

canine and bovine DAD1) (Mohorko et al., 2011). Structure predictions suggest two TMDs with both termini located in the cytosol. Functionally, DAD1 is important for OST complex stability (Sanjay et al., 1998; Roboti and High, 2012b).

TMEM258 (Ost5p)

While yeast Ost5p had long been known as an OST component (Reiss et al., 1997), its mammalian homolog was only recently identified (Parnas et al., 2015; Blomen et al., 2015). It is a 9 kDa protein with two TMH but its orientation in the membrane remains unclear. An interaction with the RPN1 homolog Ost1p has been proposed in *S. cerevisiae* (Reiss et al., 1997) but functional data on the mammalian homolog is not available to date.

OST4 (Ost4p)

Similar to OST5, Ost4p was first identified as a OST component in yeast (Karaoglu et al., 1997) and OST4 has later been confirmed as a component of the mammalian OST complexes (Kelleher and Gilmore, 2006; Dumax-Vorzet et al., 2013). It is the smallest OST subunit, with a molecular weight of only 4 kDa. Notably, this single TMH protein ($C_{\text{cyt}}\text{-}N_{\text{lum}}$ topology) is the only mammalian OST component for which high-resolution structural data is available to date (Zubkov et al., 2004; Gayen and Kang, 2011). The yeast homolog Ost4p has been implicated in stabilizing the interaction between Stt3p and Ost3p as well as regulating the incorporation of Ost3p or Ost6p respectively (Kim et al., 2003; Spirig et al., 2005).

1.5.4 The STT3A specific subunits DC2 and KCP2

STT3A-OST has been shown to incorporate two paralog specific subunits, DC2 (17 kDa) and KCP2 (Keratinocyte-associated protein 2, 14 kDa) (Shibatani et al., 2005). DC2 shows weak homology to the C-terminal domain of yeast Ost3p/Ost6p and has a predicted three-TMH topology. Noteworthy, it is lacking its homologues' N-terminal thioredoxin-like domain. KCP is a small 14 kDa protein with a predicted 4-TMH topology (Roboti and High, 2012a). It is the only OST component without any known homolog in the yeast complex. KCP2 may facilitate the glycosylation of selected substrate proteins (Roboti and High, 2012b). Functional analyses are hampered by the fact that KCP2 displays a tendency to be lost upon OST solubilization in detergent (Roboti and High, 2012b; Kelleher et al., 2003). Recent biochemical evidence indicates that DC2 and KCP2 are required for stable interaction of STT3A-OST with the PCC (Shrimal et al., 2017). However, loss of KCP2 results in a less

severe phenotype than knockout of DC2. The latter resembles a deletion mutant of the STT3A paralog of the catalytic subunit.

1.5.5 The STT3B specific oxidoreductase subunit

Instead of DC2 and KCP2, the STT3B complex associates with one of two other mammalian homologs of the yeast Ost3/6p, namely TUSC3 (tumor suppressor candidate protein 3, also known as N33, 36 kDa) or MagT1 (magnesium transporter 1, also known as implantation-associated protein - IAP, 35 kDa) (MacGrogan et al., 1996; Kelleher et al., 2003; Cherepanova et al., 2014).

In contrast to DC2, TUSC3 and MagT1 harbor a common thioredoxin fold in their luminal domain (Fetrow et al., 2001; Schulz et al., 2009) and likely acquire a four-fold membrane spanning topology (Karaoglu et al., 1995). The proteins have been proposed to compete with naturally occurring intramolecular disulfide bridge formation by forming transient mixed disulfides with a free thiol in a glycoprotein substrate. The current model suggests that this prevents these substrates from obtaining their mature, folded conformation, hence keeping them accessible for modification by OST (Schulz et al., 2009; Cherepanova et al., 2014). Furthermore, TUSC3 or MagT1 might be needed for peptide substrate recognition by STT3B-OST (Cherepanova et al., 2016).

1.6 Aims of this thesis

This dissertation aims to elucidate the mechanisms of co-translational membrane protein insertion and to characterize the coupling of translation, translocation and N-glycosylation at the mammalian endoplasmic reticulum.

The events occurring at the PCC during the biogenesis of secretory substrates are fairly well understood. In contrast, many open questions prevail with regard to the processes involved in the Sec61 mediated insertion of integral transmembrane proteins.

How do individual TMS partition into the membrane? The existing structural data show that a TMH can occupy the lateral gate of Sec61 (Gogala et al., 2014; Park et al., 2014). However, they represent post-accommodation states and cannot explain how the substrate has reached its final position. When did it get into contact with the lipid phase? Did it pass through the central pore or slide along the outer surface of Sec61?

Besides, there is no structural data on the insertion of polytopic transmembrane

proteins in eukaryotes. Do individual TMHs stay close to the channel or even form Sec61-proximal bundles prior to full release into the bilayer as it has been indicated by crosslinking studies? Where do they reside with respect to different translocon factors? And what is the exact role of the PCC and other translocon accessory factors in membrane protein topogenesis?

Moreover, since the majority of Sec61 substrates are N-glycosylated co-translationally at the ER membrane, it is important to understand how the OST cooperates with the ribosome and the PCC to accomplish this task.

Therefore, the first goal of this thesis was to establish a protocol for the purification of defined biogenesis intermediates of polytopic transmembrane proteins. Moreover, the chosen approach should be suitable to account for the additional layer of complexity which OST-involvement adds to the process for many naturally occurring substrates. Therefore, in contrast to available protocols (Voorhees et al., 2014; Voorhees and Hegde, 2016; Gogala et al., 2014) the strategy should enable recovery of OST-containing RTCs.

To that end, we designed different C-terminally truncated variants of the well-studied polytopic glycoprotein bovine opsin (see section 1.3) with a C-terminal ribosome stalling sequence. *In vitro* translation of these constructs in the presence of mammalian microsomal ER membranes was performed to obtain homogeneous and stable biogenesis intermediates at different insertion stages. A fast and mild isolation strategy was developed to allow for biochemical and structural characterization.

The second aim was to perform in depth structural analysis on the resulting complexes. For that purpose, the biogenesis intermediates were isolated biochemically and subjected to cryo-EM followed by single-particle analysis.

We set out to obtain high-resolution structures of defined steps along the co-translational insertion and N-glycosylation pathway which can provide new insights into the process of membrane protein insertion at the mammalian ER. Moreover, they can unravel the complex interplay between the molecular machineries involved in protein translation, translocation and N-glycosylation. Finally, a high-resolution structure of the mammalian OST complex can offer unprecedented insight into the architecture of this multisubunit enzyme.

Chapter 2

Materials and Methods

2.1 General

Autoclaved glassware and sterile laboratory material was used in all experiments. Buffers and stock solutions were prepared using deionized water and were sterile filtered before use. Nuclease-free H₂O was used for all *in vitro* transcription and translation reactions.

2.2 Vectors

All constructs are based on a pEXK4 vector (eurofins) coding for full-length (FL) bovine opsin. The gene was under a T7 promoter and flanked by a high translation initiation efficiency 5' UTR region as described in Beckmann et al. (Beckmann et al., 2001) and contained an N-terminal hemagglutinin-tag (HA-tag). Vectors coding for truncated opsin variants with a C-terminal ribosome stalling sequence from the human cytomegalovirus (CMV) gp48 upstream open reading frame 2 (uORF2) (Degnin et al., 1993; Bhushan et al., 2010; Matheisl et al., 2015) were generated by polymerase chain reaction (PCR, see section 2.3.1).

2.3 Molecular Cloning

2.3.1 Polymerase chain reaction (PCR)

PCR was used in order to generate vectors coding for C-terminally truncated opsin variants with a CMV stalling sequence. The forward primer (F_ pEXK4) was designed to anneal in the pEXK4 backbone downstream of the opsin FL coding sequence. Reverse primers (R_ OP96-HA, R_ OP109, R_ OP130, R_ OP164, R_ OP204) were designed such that they contain a region which overlaps with the opsin ORF

21 aa upstream of the desired final number of aa. In addition the reverse primers contained an overhang coding for the reverse complement sequence of the 21 aa CMV stalling peptide including a UGA stop-codon triplett. The OP96 construct was initially designed to include a C-terminal HA-Tag between the opsin coding sequence and the CMV stalling sequence. The tag was later replaced with the corresponding opsin aa using the primers F_ replaceHA and R_ replaceHA. All primer sequences are listed in table 2.1.

Table 2.1 – Primers used for plasmid modification. Primary sequence of the forward (F_) and reverse (R_) primers. The sequence coding for the reverse complement of the inserted CMV stalling sequence is underlined.

Name	Sequence
F_ pEXK4	CCGCTCACAATTCCACACAACATAC
R_ OP96-HA	<u>TCAGGGAGGGATATACTTGCATGTAAGCAGGCTGCT</u> <u>CAGTTTCTTGGCGGACAACACCAAGGGCTCGGC</u> GTAATCTGGGACGTCATACGGGTACTTGTGCTG- GACTGTGACGTAC
R_ OP109	<u>TCAGGGAGGGATATACTTGCATGTAAGCAGGCTGCT</u> <u>CAGTTTCTTGGCGGACAACACCAAGGGCTCGAA</u> GACCATGAAGAGGTCTGGCCACGGC
R_ OP130	<u>TCAGGGAGGGATATACTTGCATGTAAGCAGGCTGCT</u> <u>CAGTTTCTTGGCGGACAACACCAAGGGCTCGCC</u> CGTGGGCCCCAAAGACGAAGTACCC
R_ OP164	<u>TCAGGGAGGGATATACTTGCATGTAAGCAGGCTGCT</u> <u>CAGTTTCTTGGCGGACAACACCAAGGGCTCCAT</u> GGGCTTGACACACCACCGTACCG
R_ OP204	<u>TCAGGGAGGGATATACTTGCATGTAAGCAGGCTGCT</u> <u>CAGTTTCTTGGCGGACAACACCAAGGGCTCCAT</u> GCCCTCCGGGATGTACCTGGACCA
F_ replaceHA	AAGCTGCGCACACCCCTCAACTACATCGAGCCCTT GGTGTGTTGCCGC
R_ replaceHA	CTTGTGCTGGACTGTGACGTAC

The reactions were prepared using the Thermo Scientific Phusion Flash High-Fidelity PCR Master Mix according to the manufacture's protocols. PCR was performed in a two-phase protocol. Initially, reactions were prepared to contain either the forward or one of the reverse primers (F-mix and R-mix respectively). Therefore, this first phase serves for amplification of the full-length plasmid including

the primer-encoded overhangs. Subsequently, one F-mix was pooled with one R-mix and the PCR was resumed in order to obtain linearized plasmids coding for the desired opsin construct. The two step program for whole plasmid amplification is shown in detail in table 2.2.

Table 2.2 – PCR program used for plasmid mutagenesis and amplification. A two-phase program was used in order to amplify plasmids of different C-terminal truncations of bovine opsin and to introduce a C-terminal CMV stalling sequence.

Phase	Step	Purpose	Temperature	Duration	Repetitions
I	1	Initial DNA denaturation	98°C	180s	1x
	2	Iterative DNA denaturation	98°C	20s	4x
	3	Primer annealing	62°C	25s	
	4	Iterative Elongation	72°C	210s	
	5	Final Elongation	72°C	600s	1x
	6	Finish	12°C	600s	1x
Pool F-mix and R-mix					
II	1	Initial DNA denaturation	98°C	30s	1x
	2	Iterative DNA denaturation	98°C	180s	11x
	3	Primer annealing	52°C	20s	
	4	Iterative Elongation	72°C	210s	
	5	Final Elongation	72°C	600s	1x
	6	Finish	12°C	600s	1x

2.3.2 Degradation of parental DNA

After whole plasmid amplifications parental DNA was degraded using the restriction endonuclease Dpn1 (NEB). Dpn1 exclusively digests methylated DNA, hence preserving the DNA amplified by PCR. The digestion was performed at 37°C for one hour after adding 5.7 µL 10× CutSmart buffer (NEB) and 1.3 µL Dpn1 (NEB) to 40 µL of PCR product. Dpn1 treated amplified plasmids were purified using the QUIAquick PCR purification kit (QUIAGEN) according to manufacturer's instructions. DNA was eluted into 30-40 µL nuclease free water.

2.3.3 Phosphorylation and Ligation

After plasmid amplification vectors were phosphorylated and re-ligated. The Dpn1 digested PCR products were mixed in a 1:10 ratio with 10x T4 ligase buffer (NEB) and incubated for 1 h at 37°C with T4 polynucleotide kinase (NEB) in order to generate 5'phosphorylated DNA ends. Subsequently T4 ligase (NEB) was added to the reaction and the mix was incubated for 1 h at 37°C or over night at 16°C.

2.3.4 Plasmid transformation

Plasmids were transformed immediately after ligation into competent *E. coli* DH5 α cells. The cells were thawed on ice and 10 μ L of DNA were added and mixed with the cells by a quick flip of the tube. After 10 min incubation on ice the cells were heat-shocked for 90 s at 42°C in a waterbath prior to addition of 900 μ L of LB medium. Cells were recovered by incubation at 37°C for 45 min in a shaker and subsequently plated onto LB-Agar plates containing kanamycin as a selection marker (LB-Agar-Kan) for positive clones.

2.3.5 Plasmid isolation

2 mL LB medium containing 50 μ g/mL Kanamycin were inoculated with a positive clone from an LB-Agar-Kan plate. Cultures were grown for 8-9 h. Subsequently plasmid DNA was isolated using the QUIAprep Spin Miniprep Kit (QUIAGEN) according to the manufacturers instructions. Final DNA concentrations were measured using a Nanodrop 1000 device. Isolated DNA was sent for sequencing at eurofins, using the pEXK-forward and pEXK-reverse primers to confirm the primary sequence of the coding region.

2.4 Preparation of mRNA

2.4.1 Template generation by PCR

In order to generate a template for subsequent *in vitro* transcription, the region coding for the desired opsin construct was amplified from the corresponding plasmid by PCR. Used primers were a 5' T7 forward primer (TAATACGACTCACTATAG) and a 3' reverse CMV-STOP (TCAGGGAGGGATATACTTGCATG) primer. Linear DNA for subsequent use as template DNA in *in vitro* transcription reactions were amplified by PCR using the Thermo Scientific Phusion Flash High-Fidelity PCR

Master Mix and the PCR program given in table 2.3.

Table 2.3 – PCR program used amplification of opsin constructs. Linear DNA fragments coding for the T7 promoter region, the 5'UTR and the truncated opsin-ORF with a C-terminal CMV stalling sequence were amplified as templates for *in vitro* transcription.

Step	Purpose	Temperature	Duration	Repetitions
1	Initial DNA denaturation	98°C	30s	1x
2	Iterative DNA denaturation	98°C	20s	30x
3	Primer annealing	52°C	20s	
4	Iterative Elongation	72°C	210s	
5	Final Elongation	72°C	600s	1x
6	Finish	12°C	600s	1x

The products were purified using the QUIAquick PCR purification kit and concentrations were measured using a Nanodrop 1000 device. Prior to further use they were analyzed by agarose gel electrophoresis in order to confirm the correct size and rule out contaminations by secondary products. Separation of PCR fragments was done using a 2% agarose gel in TAE buffer (40 mM Tris, 20 mM acetic acid, 1mM EDTA). Agarose was resuspended in buffer and heated until it was completely dissolved. The gel was poured with 3:50000 SybrSafe DNA stain and left to polymerize for 30-45 min. For sample preparation DNA was mixed in a 1:10 ratio with 10x DNA loading dye (50% glycerol, 1 mM EDTA, 0.25% bromophenol blue, 0.25% xylene cyanol FF). A DNA molecular weight standard (100 bp or 1 kb ladder, Thermo Scientific) was loaded next to samples for size estimation. Electrophoresis was performed for 30-45 min at 125 V. Fragments were detected using a Multi Gel Jet (INTAS) system.

2.4.2 *In vitro* transcription

In vitro transcription reactions were performed in order to generate mRNA templates for *in vitro* translation. The reaction mix was prepared in the order given in table 2.4 at RT.

The mix was incubated for 1 h at 37°C, the first 15 minutes shaking slightly. Then, additional 4 µL of T7 polymerase were added and the incubation was continued for 1 h. 3 µL of TURBO DNase (from the T7 mMESSAGE mMACHINE kit, Ambion)

Table 2.4 – Reaction mix for *in vitro* transcription. The reaction mix was prepared in the given order of components at room temperature.

Component	c(stock)	V or c(final)
T7 Transcription Buffer	10x	1x
DTT	1 M	0.5 mM
ATP	125 mM	6.25 mM
GTP	125 mM	6.25 mM
CTP	125 mM	6.25 mM
UTP	125 mM	6.25 mM
Template DNA	100-300 ng/ μ L	30 ng/ μ L
Anti-RNase (Ambion)	20 U/ μ L	0.4 U/ μ L
Nuclease free H ₂ O		ad 96 μ L
T7 RNA Polymerase (pre-prepared in-house)		4 μ L
Total Volume		Σ 100 μ L

were added and the mix was incubated on a shaker for 15 min at 37°C.

For precipitation of mRNA 100 μ L of nuclease free H₂O and 120 μ L of saturated LiCl solution were added and the resulting mix was incubated over night at -80°C. The RNA was pelleted by centrifugation, the supernatant was sucked away carefully and the pellet was washed once with 70% Ethanol in order to remove salt contaminations. After a final pelleting step, the pellet was dried at RT and resuspended in 100-200 μ L nuclease free water. mRNA concentration was determined on a Nanodrop 1000 device.

Quality of the resulting mRNA was evaluated by agarose gel electrophoresis. These experiments were performed in analogy to DNA analysis experiments but samples were prepared using a 2x RNA loading dye (95% formamide, 0.025% xylene cyanol FF, 0.025% bromophenol blue, 0.025% SDS, 18 mM EDTA).

2.5 *In vitro* translation reactions

All *in vitro* translation assays were performed using the rabbit reticulocyte lysate (RRL) cell free *in vitro* translation system (Promega). *In vitro* translation was performed either in absence or presence of puromycin/high-salt treated microsomal membranes from dog pancreas (PKRMs). In absence of PKRMs, the expression

level was analysed and optimized for all constructs. Translation reactions were supplemented with PKRMs (prepared by Marko Gogala, Gene Center Munich, as described before, see Walter and Blobel, 1983) in order to evaluate the targeting and glycosylation efficiencies of the constructs used.

Before use in the reaction, mRNA was heated to 75°C for 3 min in a shaker and cooled down quickly on ice for 2 min in order to resolve any secondary structure. If PKRMs were used in the reaction, they were pre-treated with 7.5 U/μL Anti-RNase (Ambion) for 10 min on ice prior to addition to the reactions, unless specifically stated otherwise. This pre-treatment serves to minimize deleterious effects from RNases present in the microsome preparations.

Tables 2.5 and 2.6 give an overview on the reaction mix for assays without and with pre-treated PKRM, respectively.

Table 2.5 – Reaction mix for *in vitro* translation without membranes.

Component	c(stock)	V or c(final)
RRL		21 μL
Aminoacid mix	1 mM	1.2 μL
Anti-RNase (Ambion)	20 U/μL	0.4 U/μL
mRNA	1000-2000 ng/μL	100 ng/μL
Nuclease free H ₂ O		ad 30 μL
Total Volume		Σ 30 μL

Translation was performed at 28°C for 30 min in total, including the time before addition of PKRM for membrane-containing reactions. The time was adjusted to 45 min for constructs with more than 150 aa. The reactions were stopped by addition of 1 μL 1 mg/mL cycloheximide. Afterwards, each sample was immediately loaded on 100 μL sucrose cushion (30 mM HEPES/KOH pH 7.5, 10 mM Mg(OAc)₂, 100 mM KOAc, 1 mM DTT, 1 μg/mL cycloheximide, 500 mM sucrose) in a TLA 100 tube. Samples were centrifuged for 10 min at 33 k rpm/4°C in order to pellet microsomes and membrane bound ribosomes (MB). Supernatants were quickly transferred to a fresh TLA 100 tube and free ribosomes which did not bind to the microsomes (F) were pelleted for 30 min at 100 k rpm/4°C. Immediately after the spin, supernatants were quickly removed. The pellets from both spins (MB and F) were resuspended in 20 μL 1x SDS-sample buffer (1x SDS-SB; 50 mM tris(hydroxymethyl)aminomethane (Tris) / HCl pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 0.1% (w/v) bromophenol

Table 2.6 – Reaction mix for *in vitro* translation in the presence of PKRM.

Component	c(stock)	V or c(final)
RRL		21 μ L
Aminoacid mix	1 mM	1.2 μ L
Anti-RNase (Ambion)	20 U/ μ L	0.4 U/ μ L
mRNA	1000-2000 ng/ μ L	100 ng/ μ L
Nuclease free H ₂ O		ad 26.8 μ L
After 5 min incubation: pre-treated dog PKRM	0.56 eq/ μ L	0.05-0.06 eq/ μ L
Total Volume		Σ 30 μ L

blue, 100 mM 1,4-dithiothreitol). The resuspended pellets were either used immediately for analysis by SDS-PAGE and semi-dry western blotting (WB) or stored at -20°C until analysis was performed.

2.6 Protein analysis

2.6.1 Protein precipitation

For analysis by SDS-Polyacrylamide-Gel-Electrophoresis (SDS-PAGE) all protein samples with volumes larger than 25 μ L were precipitated with trichloroacetic acid (TCA). To that end samples were filled to 1 mL with H₂O and 100 μ L sodium-desoxycholate as well as 100 μ L 72% TCA were added. The precipitation mix was incubated on ice for 20 min and pelleted in a pre-cooled tabletop centrifuge for 20 min at 14 k rpm/4°C. The supernatant was sucked away carefully. The pellets were washed with 1 mL ice-cold acetone, re-pelleted under the same conditions as before and left to dry under the hood at RT. Finally they were resuspended in 15 μ L 1x SDS SB.

2.6.2 Deglycosylation by Endoglycosidase H (Endo H)

Endo H reverts N-glycosylation by hydrolyzing the chitobiose core of the oligosaccharide, leaving only a single N-acetylglucosamine residue attached to the peptide (Maley et al., 1989). For this purpose an aliquot (6-10 μ L, depending on concentration) of a membrane-bound sample was treated with Endo H (500000 U/mL, NEB).

To do so, the aliquot was mixed with 2 μ L glycoprotein denaturation buffer (NEB) and filled to 20 μ L with H₂O. The mix was incubated for 5 min at 95°C and cooled down to room temperature (RT) prior to addition of 3 μ L 10x deglycosylation buffer (NEB), 3 μ L EndoH and 4 μ L H₂O. Deglycosylation was performed for 1 h at 37°C. Subsequently the sample was frozen in liquid N₂ and stored at -20°C until used for further analysis.

2.6.3 SDS-Polyacrylamide-Gel-Electrophoresis (SDS-PAGE)

SDS-PAGE for separation of proteins according to their molecular weight was performed using standard protocols (Laemmli, 1970). Separation was performed either on 15% SDS-polyacrylamide (SDS-PAA) gels or on 12% Bis-Tris-PAA gels. All samples were prepared with 1x SDS-SB. Samples were denaturated prior to loading for 5 min at 65°C in order to preserve the peptidyl-tRNA bond. Electrophoresis was performed for 1 h at 200 V in SDS running buffer (25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS) for SDS-PAA gels and for 1 h at 180 V in MOPS running buffer (50 mM MOPS, 50 mM Tris Base pH 7.7, 0.1% SDS (w/v), 1 mM EDTA) for Bis-Tris-PAA gels.

Gels were used for western blot (WB) transfer immediately after electrophoresis. After transfer, gels were stained with Simply Blue Coomassie staining solution (Novex) to evaluate transfer efficiency. Gels were boiled in H₂O twice to remove SDS. Subsequently, gels were heated in Simply Blue staining solution and incubated on a shaker for 5-10 min at RT to visualize bands.

2.6.4 Semi-dry western blotting

Semi-dry western blotting was used in order to transfer proteins to a polydivinylfluoride (PVDF) membrane. Methanol-based blotting buffer was used (20% (v/v) methanol, 48 mM Tris/HCl, 39 mM glycine, 0.037% (w/v) SDS) and blots were performed at 75-120 mA/gel (constant current) for 1 h in a standard semi-dry blotting apparatus (BioRad).

Membranes were stained with amido-black (0.1% (w/v) naphthol blue black, 7.5% (v/v) acetic acid, 20% (v/v) ethanol) for 2 minutes on a shaker at RT and destained (40% (v/v) ethanol, 10% (v/v) acetic acid) until clear bands were visible on a light background.

Prior to antibody detection, the membranes were incubated in TBS (20 mM Tris HCl pH 7.6, 150 mM NaCl)/5% (w/v) milk-powder for 30-60 min at RT or over night

at 4°C in order to reduce unspecific binding. Subsequently the blots were rinsed with TBS. Antibody detection of the HA-tag in the nascent chain was performed with either a single-step α HA-HRP antibody (rat anti-HA-Peroxidase, High affinity 3F10, Roche) or in a two-step procedure using the α HA-probe-antibody (mouse- α HA, Santa Cruz) with a suitable secondary antibody (goat- α -mouse-HRP, Santa Cruz) according to the following protocols:

- α HA-HRP
 1. incubate 1 h at RT in TBS/5% (w/v) milk-powder, 1:5000 α HA-HRP
 2. wash 3×10 min with TBS-T (20 mM Tris HCl pH 7.6, 150 mM NaCl, 0.1% (v/v) Tween)
- α HA-probe
 1. incubate over night at 4°C in TBS/2% Bovine serum albumin, 1:1000 α HA-probe
 2. wash 3×10 min: TBS-T, TBS, TBS
 3. block for 10 min at RT with TBS/5% (w/v) milk-powder
 4. incubate 1 h at RT in TBS/5% (w/v) milk-powder, 1:5000 goat- α mouse-HRP
 5. wash 3×10 min with TBS-T

For detection, the membranes were incubated with ECL substrate (Thermo Fisher) or super signal ECL substrate (Thermo Fisher). Chemiluminescence was visualized either on LucentBlue X-ray film (advantia) which was fixed using a developing device (Optimax Type TR, PROTEC) or alternatively by using a AI600 imaging device (GE Healthcare). Exposure times were adjusted according to signal strength.

2.6.5 Mass-spectrometry

In order to learn about the protein composition of the OP96-cryo-EM sample, it was subjected to mass spectrometry (MS) analysis. Therefore, the sample was subjected to SDS-PAGE and excised from the gel. Gel extraction, proteolytic digest, MS and data analysis was performed by Thomas Fröhlich (Gene Center, LMU). Proteins were identified by screening against peptide libraries from *canis lupus familiaris* (dog) for membrane components and libraries from *oryctolagus cuniculus* (rabbit) for cytosolic components. Hits with very high (≥ 100) and high (≥ 30) confidence scores

were combined and rated according to abundance. To that end, the exponentially modified protein abundance index (emPAI) was used as a label free quantitative measure to estimate relative protein amounts in the sample (Ishihama et al., 2005).

2.7 Purification of opsin biogenesis intermediates for cryo-EM

The purification strategy was optimized to preserve the very transient OST-containing RTC intermediates. Hence, the final protocol did not contain an affinity purification step via the nascent chain in contrast to previous studies (Gogala et al., 2014; Voorhees and Hegde, 2016). This greatly reduces the time window between *in vitro* translation and vitrification and minimizes shearing forces on the complexes of interest. The resulting increase in sample heterogeneity was compensated for by extensive classification during cryo-EM data analysis (see section 2.8).

Buffers

The following buffers were used during the purification:

- Compensation Buffer (CB): 150 mM HEPES/KOH pH 7.5, 50 mM Mg(OAc)₂, 2.5 M KOAc, 5 mM DTT, 5 µg/mL cycloheximide
- High-Salt Sucrose Cushion (HS-Suc): 30 mM HEPES/KOH pH 7.5, 10 mM Mg(OAc)₂, 500 mM KOAc, 1 mM DTT, 1 µg/mL cycloheximide, 500 mM Sucrose
- Ribosome buffer (RB): 30 mM HEPES/KOH pH 7.5, 10 mM Mg(OAc)₂, 100 mM KOAc, 1 mM DTT, 1 µg/mL cycloheximide
- Solubilization Buffer (SB): 30 mM HEPES/KOH pH 7.5, 10 mM Mg(OAc)₂, 100 mM KOAc, 1 mM DTT, 1 µg/mL cycloheximide, 1.5 % Digitonin
- Digitonin Sucrose Cushion (DSuc): 30 mM HEPES/KOH pH 7.5, 10 mM Mg(OAc)₂, 100 mM KOAc, 1 mM DTT, 1 µg/mL cycloheximide, 500 mM Sucrose, 0.3 % Digitonin
- Digitonin Grid Buffer (DGB): 30 mM HEPES/KOH pH 7.5, 10 mM Mg(OAc)₂, 100 mM KOAc, 1 mM DTT, 1 µg/mL cycloheximide, 0.3 % Digitonin

***In vitro* translation and high-salt treatment**

Six 50 μ L *in vitro* translation reactions with PKRMs were prepared as described in section 2.5 including pretreatment of PKRM and denaturation of mRNA secondary structure.

After stopping each reaction with 1 μ L cycloheximide (c=1 mg/mL), 12.5 μ L CB was added to each reaction and the resulting mix was incubated for 30 min at 25°C, shaking slightly. For the OP204 sample, the addition of cycloheximide was omitted and *in vitro* translation was immediately followed by incubation under high-salt conditions.

Solubilization and isolation of membrane bound ribosomal complexes

Unless stated otherwise, all resuspension steps were performed on ice and all centrifugation steps were carried out at 4°C.

All samples were pooled and applied to 600 μ L HS-Suc in a TLA120.2 tube. Pelleting of membranes was performed for 10 min at 45 k rpm (spin 1). The supernatant (SN1) was removed immediately and transferred to a fresh TLA120.2 tube. The membrane pellet was rinsed immediately with 100 μ L RB (saved for analysis = W) and 400 μ L SB were added. SN1 was spun at 100 k rpm for 45 min in order to pellet free ribosomes, which had not bound to the PKRM (spin 2). The supernatant from spin 2 was quickly discarded and the ribosome pellet was carefully resuspended in 60 μ L RB over 2 h by pipetting up and down every 15 min. Meanwhile, solubilization of the membrane pellet was carried out for 1 h by carefully mixing the sample by pipetting every 10 min. The resulting solubilization mix was loaded on 600 μ L DSuc in a fresh TLA120.2 tube and spun at 100 k rpm for 45 min to pellet solubilized RTCs (spin 3). The supernatant (SN2) was quickly removed and saved for analysis and the pellet was resuspended in 60 μ L DGB for 90 min, carefully pipetting up and down every 10 min. The solubilized sample was transferred to an eppendorf tube and spun for 30 s in a tabletop centrifuge to remove large aggregates. The supernatant was transferred to a fresh tube and adjusted to a final concentration of 5.0-5.5 A₂₆₀/mL with DGB. An overview on the sample preparation is given in figure 3.4 of section 3.2. Aliquots were taken at different steps during the purification and analyzed by SDS-PAGE and WB.

2.8 Cryo-EM data collection

Cryo-EM data collection

Carbon coated holey grids (2 nm, R3/3, Quantifoil) were glow discharged at 0.2 mbar for 30 s. For each grid, 3.5 μ L of sample were applied to the grids at a concentration of 5.0-5.5 A260/mL. Subsequent vitrification was performed by plunge freezing in liquid ethane using a Vitrobot mark IV (FEI Company, Netherlands) with a blotting time of 3 s at 4°C. Cryo-EM data was collected semi-automatically using the acquisition software EM-TOOLS (TVIPS, Germany) on a Titan Krios transmission electron microscope (FEI Company) at a defocus range between 0.9 and 3.5 μ m. All data were recorded on a Falcon II detector upgraded with a Falcon III detector chip under low dose conditions with a nominal pixel size of 1.084 Å/pixel (px) on the object scale. All micrographs experienced a total exposure of 28 electrons/Å² fractionated into 10 frames. Grid preparations were performed by Susanne Rieder and cryo-EM data collection was performed by Dr. Otto Berninghausen.

In total five different cryo-EM datasets were collected from three different samples: OP96 (dataset I, 7120 micrographs; dataset IV, 13087 micrographs), OP109 (dataset II, 9758 micrographs; dataset V, 12656 micrographs) and OP204 (dataset III, 8563 micrographs).

Contrast transfer function (CTF) estimation was performed on-the-fly with CTFFIND4 (Mindell and Grigorieff, 2003; Rohou and Grigorieff, 2015) and micrographs with an estimated resolution limit of below 4.5 Å were automatically discarded.

Original image stacks were summed up and aligned at micrograph level using MotionCor2 (Zheng et al., 2017).

Cryo-EM data processing

The CTF parameters were re-estimated via GCTF (Zhang, 2016). All micrographs were screened manually for ice quality prior to automated particle picking with Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang>). All classifications and refinements were performed using Relion-2.1 (Kimanius et al., 2016).

Two distinct classification routines were used. Initial structural models from datasets I-III (see section 3.3) were obtained by the following classification scheme:

1. 2D classification
2. selection of good classes with ribosomal particles
3. initial 3D refinement with an undecimated box size of 420 px, resized to 100 px
4. 3D classification, 10 classes

5. Optional (for partial OST occupancy): 3D classification of P-site peptidyl-tRNA class
6. 3D refinement of P-site peptidyl-tRNA containing class with an undecimated box size of 420 px
7. post-processing

Data obtained by extended data collection for OP96 was processed according to a slightly modified routine. The two datasets collected for the OP96 sample (I and IV) were joined after automated particle picking and reference-free 2D classification resulting in 1.343.416 ribosomal particles. After an initial round of 3D refinement these were subjected to extensive 3D classification. Differences between the resulting eleven classes comprised the presence or absence of the SSU, OST and eEF2 as well as different tRNA states. Also conformational differences in the L1 stalk and the relative arrangement of ribosomal subunits were found. A clean class with densities for both P-site peptidyl-tRNA and OST (188.900 particles; 14.1 %) was refined to high resolution (reconstruction A1) using a box size of 550 px, resized to 500 px. Two further classes (13.6 %) were lacking density for tRNAs and instead had strong occupancy of eEF2 as well as partial occupancy of the OST complex. These were subjected to another round of 3D classification using a mask on the translocon region where they clearly separated into populations with or without OST. OST-containing particles (90.895 particles, box size 550 px, resized to 500 px) were refined to high resolution (reconstruction B1). The final refinement was continued for reconstructions A1 and B1 including a mask on the ligand region, resulting in reconstructions A2 and B2. All final reconstructions were subjected to post-processing using a wide soft-edge mask. This resulted in final resolutions of 3.2 Å (Reconstruction A1), 4.2 Å (Reconstruction A2), 3.3 Å (Reconstruction B1) and 4.7 Å (Reconstruction B2) according to the FSC 0.143 criterion following the Relion gold-standard refinement. Local resolution was calculated using the ResMap option implemented in Relion. Data from the second collection of the OP109 sample (dataset V) was processed by Robert Buschauer (Gene Center, LMU Munich) identically to the extended OP96 routine but without combining it with the first OP109 collection.

2.9 Cryo-EM data analysis and model building

The post-processed maps were low-pass filtered at several defined resolutions for model building and refinement in order to account for differences in local resolution. All existing pdb structures used as starting points for model building were docked

into the maps as a rigid body using UCSF Chimera (Pettersen et al., 2004). Since the programmed population displayed higher resolution than the corresponding non-programmed map both before and after masked refinement, manual adjustments to pdbs were performed in Coot based on the A1 and A2 maps respectively (Emsley and Cowtan, 2004).

Models for the LSU and the P-site tRNA (pdb 3jah, Brown et al., 2015) were fitted into the A1 map. The region contacting the cytosolic RPN1 domain was manually adjusted by Dr. Thomas Becker (Gene Center, LMU Munich). Namely, the C-terminus of eL28 was remodeled and extended by 11 aa and the rRNA H19/H20 was rearranged to fit the electron density.

Sec61 was modeled in the A2 map based on the structure of the 'primed' state (pdb 3j7q, Voorhees et al., 2014). Minor adjustments were made in the plug region (res. 58-66) and TMH10 (res. 436-465). The hinge region (res. 196-242) as well as the N-terminus (res. 1-27) were modeled *de novo*.

Homology models (HMs) for the STT3A TM and luminal domains were created using the PHYRE2 server (Kelly et al., 2015) based on the crystal structure of the archaeal homolog AglB (pdb 3waj, Matsumoto et al., 2013) by Thomas Becker (Gene Center, LMU Munich). The initial models for the TM and soluble domains were combined and unambiguously placed in the OST density of the locally refined A2 map. The resulting HM was manually adjusted in Coot based on electron density features and the recent high resolution yeast OST structure (pdb 6ezn, Wild et al., 2018).

The TMHs of the non-catalytic OST-subunits were initially assigned based on the following criteria:

- predicted number of TMHs for different subunits using the jPred4 secondary structure prediction (Drozdetskiy et al., 2015) as well as the ΔG prediction server (Hessa et al., 2007)
- connectivity of electron densities between putative TMHs
- electron densities for bulky sidechains
- biochemical prediction of subcomplexes

The assignment was later confirmed by the high resolution cryo-EM structures of the *S. cerevisiae* OST complex (see section 4.2) (Wild et al., 2018; Bai et al., 2018).

The cytosolic four-helix bundle of RPN1 which does not form part of the *S. cerevisiae* complex was assigned by Dr. Stefan Pfeffer (Max Planck Institute for Biochemistry, Martinsried) based on secondary structure prediction using the jPred4 server.

For the well-resolved metazoan-specific DC2 subunit (local resolution ranging from

3.5-4.2 Å) an atomic model of the three TMH was built based on the jPred4 secondary structure prediction and positioning of bulky side chains. Homology models for the OST4 and RPN1 TMDs were created using the PHYRE2 server based on the yeast OST structure (Wild et al., 2018) and manually adjusted in Coot. When local resolution did not allow to build an atomic model, namely for the TMHs of TMEM258 and Subcomplex III (DAD1, OST48, RPN2), for STT3A-TM9 as well as the loop regions in DC2, poly-alanine segments were placed instead. The entire translocon model was refined using PHENIX (Afonine et al., 2012).

The final models of the LSU, Sec61 and the OST complex were fitted as individual rigid bodies into the maps of the non-programmed population (B1 and B2). Moreover, they were docked individually into an improved cryo-electron tomography structure of the native OST-containing RTC (Dr. Stefan Pfeffer, Max Planck Institute, Martinsried). Since this structure contains Sec61 with an open lateral gate, the N-terminal region of Sec61 α was adjusted separately. Vectors representing the conformational rearrangements among the three resulting translocon states (see figure 3.18) were visualized using PyMol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.).

Chapter 3

Results

3.1 Expression of C-terminally truncated variants of bovine opsin

In vitro translation in presence of rough microsomal membranes (RM) or puromycin/high-salt treated microsomal membranes (PKRM) from dog pancreas has a long standing tradition as a model system for studying membrane protein biogenesis (Blobel and Dobberstein, 1975; Walter and Blobel, 1983). In addition, it has been shown to represent a suitable system for generation of defined translation/translocation intermediates for structural studies (Gogala et al., 2014; Voorhees and Hegde, 2016). We chose the rabbit reticulocyte lysate (RRL) cell-free *in vitro* translation system in combination with dog PKRM to produce defined intermediates of bovine opsin. Conditions for the use of non-pre-treated dog PKRMs in combination with the RRL translation system were tested with a construct coding for an N-terminal HA-tag, the first 75 opsin aa as well as a C-terminal CMV stalling sequence (see figure 3.1, panels A and B, from now on referred to as OP96). The CMV stalling peptide was introduced to stall ribosomes at a specific point with a minimal extent of subsequent nascent chain (NC) release, thereby allowing enrichment of defined ribosome nascent chain complexes (RNCs).

Expression in absence of PKRM was performed as a negative control for glycosylation and to confirm that the conditions used for microsome pelleting do not result in pelleting of ribosome species which are not bound to the membranes. Results were analyzed by SDS-PAGE and WB (see figure 3.1, panel C).

In general two different types of signals can be expected for WB detection of a NC after *in vitro* translation. First, there should be a signal corresponding to the peptidyl-tRNA representing intact tRNA-NC species. However, partial hydrolysis of the peptidyl-tRNA bond usually leads to a second population of free peptide which can be detected at lower molecular weight (MW). For samples which have

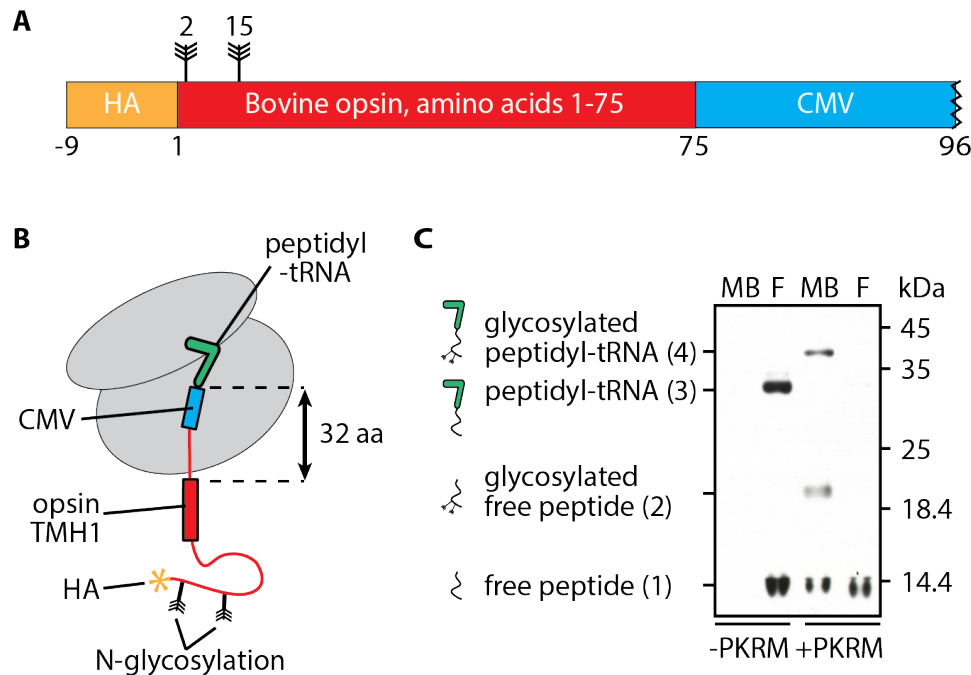


Figure 3.1 – Construct design and test expression of OP96 in absence or presence of PKRM. A) Schematic depiction of the OP96 construct used for expression tests, N-glycosylation sites are marked with fletched arrowheads B) Illustration of the expected architecture of the glycosylated OP96-RNC-complexes C) Anti-HA WB detecting NC species in the membrane bound (MB) and free (F) ribosome populations after translation without (-PKRM) or with (+PKRM) dog PKRM, samples correspond to 1/2 of the products of a 30 μ L *in vitro* translation reaction (Gel: 12% Bis-Tris; WB: PVDF, 75 mA/gel, 1 h; Antibody: α HA-probe/goat- α -mouse-HRP; Detection: Super ECL, film, 1 min exposure).

been translated in presence of PKRM one can furthermore expect glycosylated variants of both NC populations which experience a shift to higher MW due to the modification as soon as the RNCs have been successfully targeted to the membranes and engaged the translocation and glycosylation machineries. Here, when *in vitro* translation was performed in absence of PKRM, chemiluminescence signal on the anti-HA WB could only be detected in the free ribosome fraction but not in the pellet fraction of the microsome-pelleting step. The two signals visible in the free ribosome pellet correspond to the non-glycosylated peptidyl-tRNA and non-glycosylated free peptide, respectively. Partial hydrolysis of the peptidyl-tRNA bond is due to the basic pH of the 1x SDS-SB. Upon addition of PKRM, both signals could be detected in the membrane-bound fraction and experienced a shift to higher MW, indicating successful glycosylation. Non-glycosylated free peptide in the membrane bound sample could either be a contamination from the supernatant or represent NCs which failed

to be glycosylated and were released from the membrane-bound ribosomes. Signal for the free peptide was also detected in the free ribosome pellet of the PKRM-containing reaction, indicating ribosome species which failed to be targeted to the membranes and lost their NC during analysis. The experiment demonstrates that the chosen centrifugation conditions are suitable for selective isolation of membrane bound ribosomes.

The efficiency of targeting and glycosylation was estimated by comparing intensities of the corresponding signals. For this specific test experiment targeting was estimated to 60% (MB vs. F signal in +PKRM sample). Glycosylation efficiency was estimated by comparing glycosylated and non-glycosylated species in the +PKRM/MB fraction. Assuming that the non-glycosylated free peptide is not due to a supernatant contamination but rather due to incomplete modification, glycosylation efficiency amounted to approximately 50%. Finally, it became apparent that the overall translation efficiency is decreased in the presence of PKRM (Total signals -PKRM vs. +PKRM). This was due to contaminations of microsome preparations with endogenous RNase species.

A recent study by Vermeire *et al.* suggests that pre-treating PKRMs of different species with RNase inhibitor can reduce these deleterious effects (Vermeire *et al.*, 2015). Also, it was in principle desirable to increase the amount of PKRM present in the *in vitro* translation reaction to obtain higher targeting and glycosylation efficiencies. Therefore, two different amounts of dog PKRM were tested in *in vitro* translation assays of OP96 with and without anti-RNase pre-treatment prior to use in the experiment. The efficiency of targeting and glycosylation was analyzed by SDS-PAGE and WB (see figure 3.2).

In contrast to the test expression analyzed in figure 3.1, no signal was present in the free ribosome fraction indicating highly efficient targeting. However all reactions showed signals representing the glycosylated forms of peptidyl-tRNA and free peptide in the membrane-bound populations. The fact that no additional signal was detected at the MW of the non-glycosylated species suggested a very high glycosylation efficiency. Moreover, it could be seen that pre-treatment increases the efficiency for both PKRM concentrations. The strongest signals were observed for the reaction with 0.06 eq PKRM/ μ L (equal to 1.8 eq PKRM in a 30 μ L translation reaction) after anti RNase pre-treatment. Therefore, these conditions were used for all subsequent translation experiments.

The next step was to expand the analysis to opsin constructs of varying NC lengths in order to mimic different stages of biogenesis. Truncation sites were chosen such

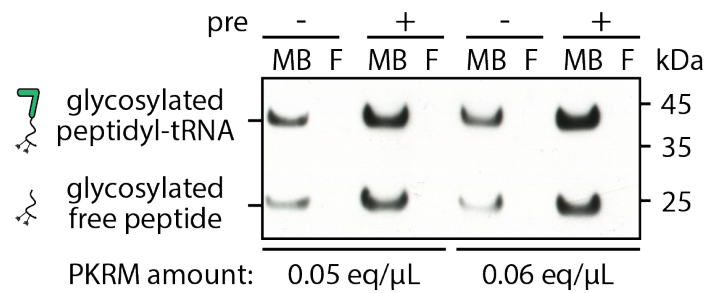


Figure 3.2 – Optimizing PKRM for use in *in vitro* translation. Anti-HA WB detecting NC species in the membrane bound (MB) and free (F) ribosome populations, "pre" indicates PKRM pre-treatment with anti RNase prior to use in the assay, samples correspond to 1/2 of the products of a 30 μL *in vitro* translation reaction and PKRM amounts are indicated per μL reaction (Gel: 12% Bis-Tris; WB: PVDF, 75 mA/gel, 1 h; Antibody: αHA-probe/goat-α-mouse-HRP; Detection: Super ECL, film, 8 min exposure).

that - without the N-terminal HA-tag - the NC including the CMV stalling sequence had the same number of aa as constructs used for existing crosslinking studies on opsin biogenesis (Laird and High, 1997; Meacock et al., 2002; Ismail et al., 2006; Ismail et al., 2008). Since these studies have revealed an increasing number of incompletely translated products for long NC lengths (Ismail et al., 2008), constructs coding for more than the N-terminal four opsin TMS were excluded from the analysis. Opsin-CMV constructs with NC lengths of 96, 109, 130, 164 and 204 aa were expressed in absence and presence of pre-treated dog PKRMs. Reactions without PKRMs were performed for each construct as a positive control for translation and as a control for the MW of the non-glycosylated NC. As an additional control for each construct, one reaction was performed in presence of PKRM and the MB pellet was subsequently treated with Endo H to revert glycosylation. Analysis was performed by SDS-PAGE and WB and is visualized in figure 3.3.

All constructs were expressed *in vitro* and were targeted and glycosylated successfully in the presence of pre-treated dog PKRMs. However, the expression efficiency varied considerably when comparing different constructs. Since all constructs had an identical 5'UTR and start codon environments it was unlikely that they displayed entirely different expression efficiencies. Instead, it had been observed that translation efficiency for one construct can vary significantly between mRNA preparations which was likely causing these differences.

For the OP204 construct, the signal for the PKRM containing samples was significantly reduced. Since the same batch of PKRM was used for all reactions, this variation is rather due to sample handling than actual negative effects of the PKRMs.

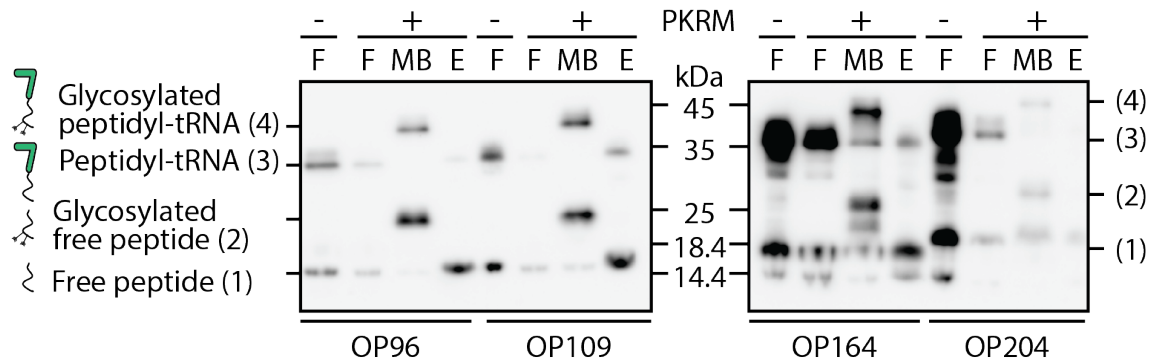


Figure 3.3 – Test expression of different opsin NC lengths in absence or presence of PKRM. Anti-HA WB detecting NC species for the OP96, OP109, OP130, OP164 and OP204 constructs in the membrane bound (MB) and free (F) ribosome populations after translation without (-) or with (+) dog PKRM and including a MB sample after EndoH treatment (E), samples correspond to 1/2 of the products of a 30 μ L *in vitro* translation reaction (Gel: 12% Bis-Tris; WB: PVDF, 120 mA/gel, 1 h; Antibody: α HA-HRP; Detection: ECL, Imager, 3 min exposure).

3.2 Preparation of solubilized opsin biogenesis intermediates

The OP96, OP109 and OP204 constructs were chosen for analysis by cryo-EM. The OP96 construct represents a very early intermediate and was selected to gain insights into the initial Sec61 gating process. OP109 reflects a NC length, which should enable full accommodation of the signal anchor-helix (SA-helix). The OP204 construct was chosen based on the findings that the TMHs 1-3 constitute an individually folding opsin subdomain. The construct was therefore considered suitable to challenge the hypothesis of TMH retention and Sec61-proximal bundling.

The preparation of samples for cryo-EM analysis is described in detail in the methods section (section 2.7) and is visualized in figure 3.4. In summary, after *in vitro* translation, samples were subjected to a high-salt (HS) incubation prior to pelleting of PKRM and solubilization of membrane bound ribosomal complexes using digitonin. Finally, the solubilized ribosome-translocon complexes (RTCs) were purified and concentrated by another pelleting step and resuspended in detergent-containing grid buffer (DGB). In contrast to previously established protocols, the preparation does not include an affinity purification via the nascent chain, thereby minimizing sample preparation time, in order to preserve fragile assemblies. The cryo-EM preparations of the OP96, OP109 and OP204 constructs were analyzed by SDS-PAGE and WB. For simplicity, only the lanes giving signal

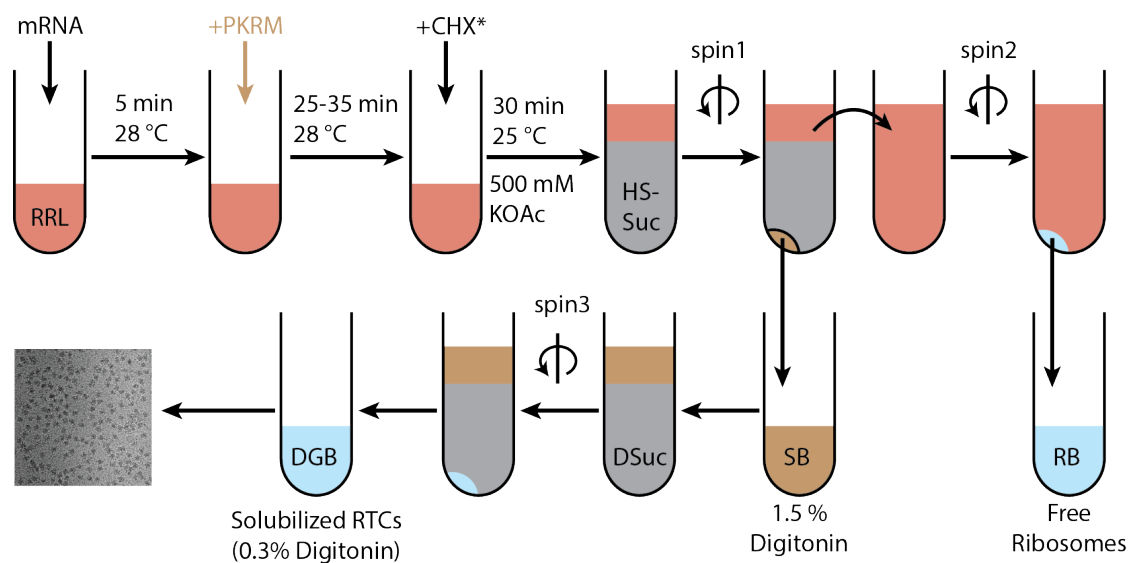


Figure 3.4 – Cryo-EM sample preparation. Schematic depiction of the sample preparation for cryo-EM analysis of the OP96, OP109 and OP204 constructs. CHX = cycloheximide (* omitted for the OP204 sample), HS-Suc = high-salt sucrose cushion, SB = solubilization buffer, D-Suc = digitonin sucrose cushion, DGB = digitonin grid buffer, RB = ribosome buffer.

in the anti-HA WB are depicted in figure 3.5.

All final solubilized samples showed a strong signal for the glycosylated NC species indicating successful isolation of defined biogenesis intermediates.

For OP96, glycosylated free peptide which has lost its tRNA during the preparation was detected in the wash fraction after membrane pelleting (W) and the supernatant after solubilization (SN2). In addition, minor amounts of non-glycosylated species, which have failed to stably bind the PKRMs were detected in the free ribosome pellet.

The anti-HA-WB from the OP109 and OP204 preparations also showed bands at the expected molecular weight for the final cryo-EM sample (OP109), or the pelleted free ribosomes and the final sample (OP204). However, signal could not be detected in any other fractions. This was likely due to differences among the three analyses with respect to gel type (15% SDS-PAA vs. 12% Bis-Tris-PAA), blotting current (75-120 mA/gel), choice of antibody (α HA-probe vs. α HA-HRP) as well as the detection method (film vs. imaging device). Conditions used are indicated in the figure legend.

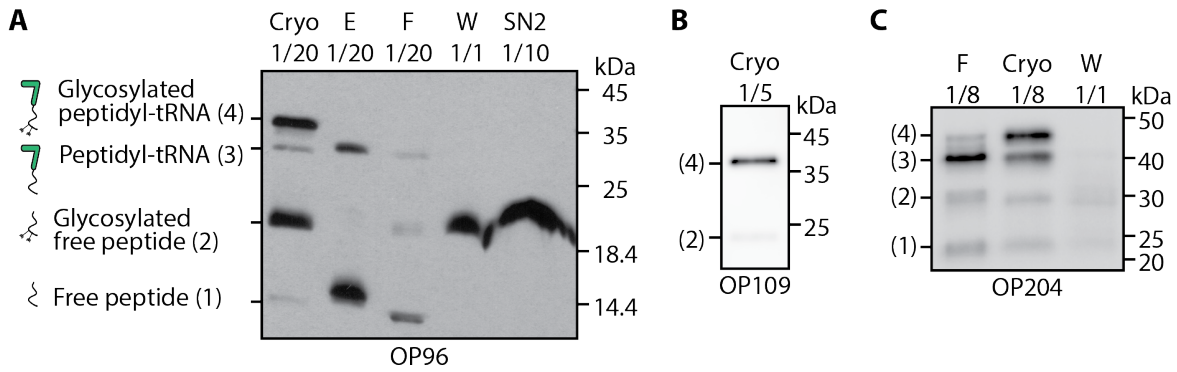


Figure 3.5 – Purification of solubilized opsin biogenesis intermediates for cryo-EM. Anti-HA WB detecting NC species in different fractions during the preparation of solubilized RTCs. (A) OP96 construct, Cryo = Final cryo-EM sample, E = Endo H-treated aliquot of final sample, F = Free ribosomes pelleted from the supernatant after membrane pelleting (spin 1, see section 2.7), W = membrane pellet wash, SN2 = supernatant after pelleting of solubilized complexes (spin 3); (Gel: 15% SDS; WB: PVDF, 100 mA/gel, 1 h; Antibody: α HA-HRP; Detection: Super ECL, film, 10 s exposure) (B) OP109 construct (Gel: 12% Bis-Tris; WB: PVDF, 120 mA/gel, 1 h; Antibody: α HA-probe/goat- α -mouse-HRP; Detection: Super ECL, Imager, 1 min exposure) (C) OP204 construct (Gel: 12% Bis-Tris; WB: PVDF, 120 mA/gel, 1 h; Antibody: α HA-HRP; Detection: ECL, Imager, 5 min exposure). A clear signal for peptidyl-tRNA is visible for all cryo-EM samples.

3.3 Preliminary cryo-EM structures of opsin biogenesis intermediates

Initial cryo-EM datasets of the OP96, OP109 and OP204 constructs were collected (datasets I-III) and processed as described in section 2.8.

In silico 3D classification revealed considerable sample heterogeneity for all constructs. This was expected due to the lack of an affinity purification step or a sucrose gradient separation of different ribosomal species in the preparation protocol. While the precise class distributions varied among the constructs, the general populations observed in the preparations were similar. All classes contained either 80S or 60S ribosomes with close-to-complete occupancy of Sec61. In addition, some classes had extra density for several translation- or translocon factors like eEF2, eIF5A, eIF6, different tRNA species, OST and TRAP. eIF6 was exclusively observed on the 60S particles, likely representing intermediates of ribosomal quality control pathways which are initiated upon ribosome stalling at the ER membrane. It is known from previous studies that eEF2, which is functionally required during the translocation step of translation elongation, also has a high affinity to non-translating ribosomes (Voorhees et al., 2014; Liu and Qian, 2016). In line with these results, presence of

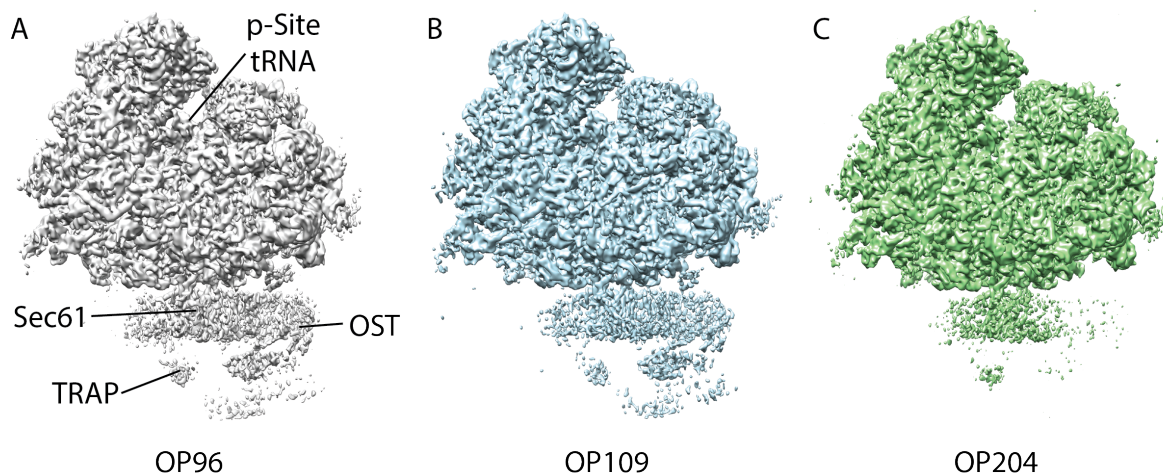


Figure 3.6 – Preliminary cryo-EM structures of three opsin biogenesis intermediates. Cryo-EM structures of the programmed solubilized RTCs programmed with three different opsin NC lengths, low-pass filtered to 5.5 Å.

eEF2 was mutually exclusive with tRNA binding in the A- or P-site, but was in some classes observed together with eIF5A (OP96, OP109) or E-site tRNA (OP204). eIF5A has previously been shown to be enriched on pull-out samples containing cycloheximide (Schmidt et al., 2015), therefore it is not surprising that it could not be detected in the OP204 sample which was prepared without the antibiotic. Notably, it was also absent from all classes containing a P-site peptidyl-tRNA in the cycloheximide-containing OP96 and OP109 preparations. Binding of accessory translocon factors appeared to be independent of translation factor occupancy. TRAP appeared delocalized or substoichiometrically bound in all classes. In contrast, the majority of otherwise identical populations (60S-Sec61-eIF6, 80S-P-Site-NC, 80S-eEF2-±eIF5A) separated into two distinct classes with or without OST.

Since our main goal was to investigate membrane insertion and N-glycosylation of the opsin model substrate, we limited our analysis of each cryo-EM dataset to the class containing a peptidyl-tRNA in the P-site and OST. These classes were refined to highest possible resolution.

The final resolutions according to the gold-standard criterion employed by the Relion 2.1 post-processing routine were 3.7 Å (OP96), 3.5 Å (OP109) and 4.1 Å (OP204) respectively. They all showed well-defined electron density for the programmed ribosome (see figure 3.6). However, the signal decreased rapidly in the Sec61 and surrounding membrane region and the translocon components could only be visualized at lower contour levels and after low-pass filtering. Therefore, these initial reconstructions only allowed for limited interpretation.

Surprisingly, the PCC conformation appeared identical for all three opsin constructs

and most closely resembled the primed state (Voorhees et al., 2014) when compared to existing structural models (see figure 3.7). In this state the lateral gate of Sec61 α is almost entirely closed, with just a slight opening at the cytosolic face. Moreover, in contrast to other structures of Sec61 in an inserting state (Gogala et al., 2014; Voorhees and Hegde, 2016), no additional α -helical densities corresponding to an inserting substrate TMH were detected in the lateral gate region in the opsin biogenesis intermediates.

These findings were contrasting our expectations, which included varying degrees

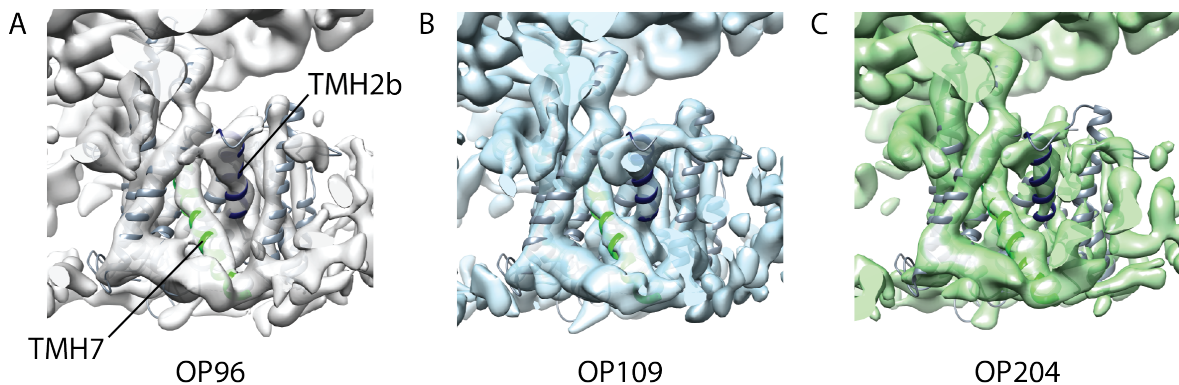


Figure 3.7 – Sec61 conformation in the solubilized opsin biogenesis intermediates.

Zoom onto the Sec61 lateral gate region in the preliminary cryo-EM structures of the OP96, OP109 and OP204 biogenesis intermediates, low-pass filtered to 7 Å. A model of Sec61 in the primed state (PDB 3j7q, Voorhees et al., 2014) with the lateral gate forming helices TMH2b and TMH7 colored in dark blue and green provides a good overall fit for all three constructs.

of lateral gate opening as well as extra densities representing different numbers of opsin TMH in proximity of the lateral gate based on available crosslinking data (Laird and High, 1997; Meacock et al., 2002; Ismail et al., 2006; Ismail et al., 2008). The lack of NC density in the lateral gate region can be explained by different scenarios:

- Location of TMHs in vicinity of the lateral gate, but low signal-to-noise ratio due to conformational and/or positional flexibility of the TMHs
- Release of the TMHs into the lipid bilayer
- Stable positioning of TMHs at a position different from the lateral gate environment

Since the preliminary reconstructions remained inconclusive regarding the membrane insertion process, we then focused on the RTC-OST interactions. The OP96

construct had the highest occupancy for a P-site peptidyl-tRNA and OST containing class (55786 particles, 17.3% of ribosomal particles). For OP109 this class represented only 12.0% of ribosomal particles (25275 particles). The OP204 dataset did not yield a clean class with high occupancy of both peptidyl-tRNA and OST but instead showed only partial OST occupancy in the final class (32802 particles, 17% of ribosomal particles). Unfortunately, this class could not be sub-classified any further due to low particle numbers. For all constructs, the density in the OST region was too weak to draw conclusions on the architecture of the complex or its interactions with Sec61 and the ribosome.

Taken together, the preliminary cryo-EM reconstructions allowed only very limited conclusions regarding the Sec61 functional state and the integration of OST into the RTC. Hence, in order to improve the signal-to-noise ratio and potentially enable further *in silico* classification of different conformational states we decided to collect more data on the most promising sample. The OP96 construct was prioritized for extended analysis because it had the highest percentage of OST-bound translating ribosomes.

3.4 Cryo-EM analysis of an early opsin biogenesis intermediate

Classification of the combined OP96 data (datasets I and IV) confirmed the high degree of heterogeneity observed for the preliminary reconstruction. The data contained both, programmed and non-programmed 80S populations as well as two translocon-bound 60S classes with and without extra density for OST. The processing scheme was focused on isolation of OST-containing RTCs harboring 80S ribosomes and led to two final classes (A and B, see figure 3.8).

Class A was programmed with a P-site tRNA and showed clear density for the NC in the ribosomal tunnel whereas class B did not contain any tRNA populations but was instead bound to the translational GTPase eEF2.

When employing the unsupervised refinement approach implemented in Relion, the signal for the translocon components was still considerably weaker than the ribosome signal (Classes A1 and B1). Besides, in contrast to Sec61 and OST, TRAP appeared to be disordered or bound in substoichiometric amounts as already observed in the preliminary reconstruction. By using a mask around the ligand region during refinement, the alignment was now based on the signal within the mask rather than on the otherwise dominating signal from the rRNA. This

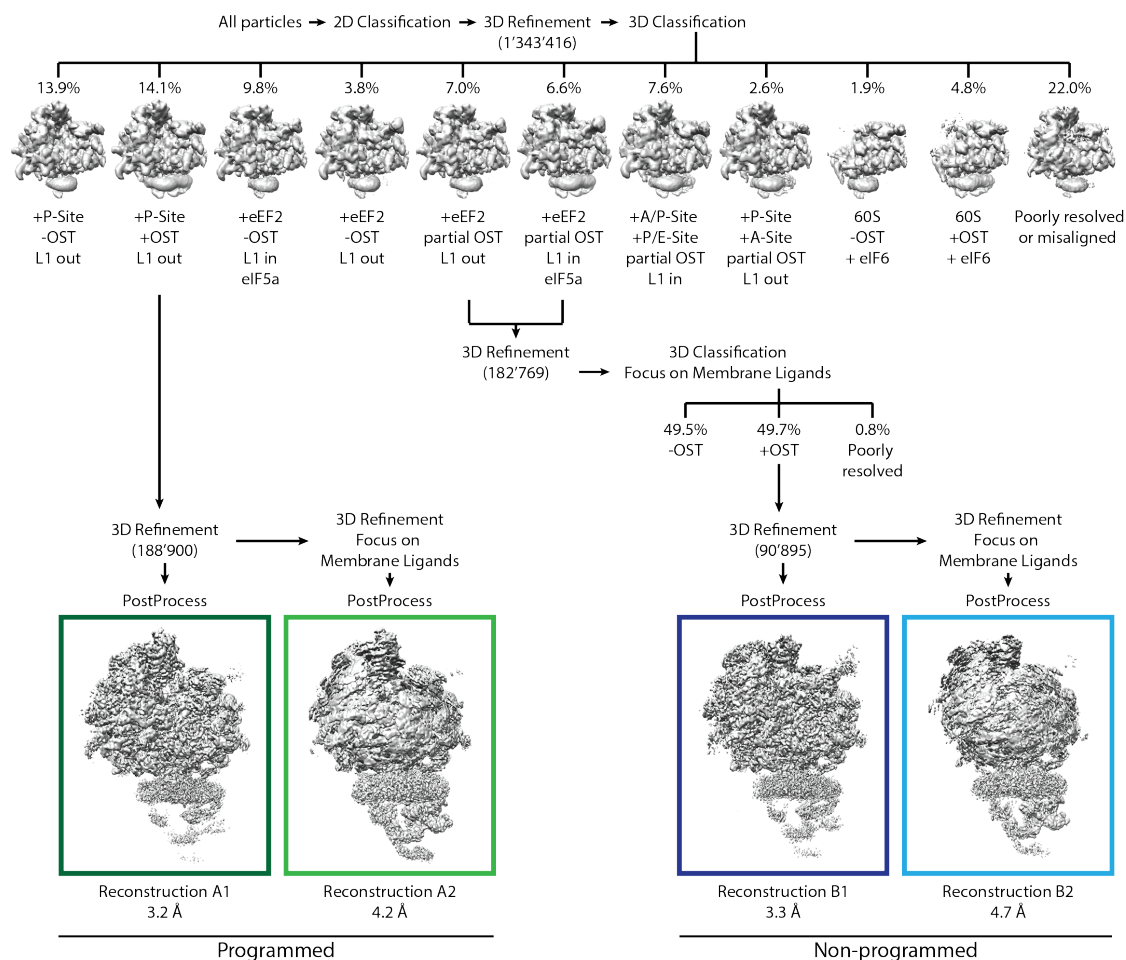


Figure 3.8 – Classification scheme of the combined cryo-EM datasets from the solubilized OP96 biogenesis intermediate. Non-ribosomal particles were excluded after 2D classification. Well-resolved classes after initial 3D classification represent eight different translocon bound 80S classes and two translocon bound 60S classes. The class with strong density for both, P-site tRNA and OST was selected (14.1 %) and refined to high resolution (Reconstruction A1). The two 80S classes lacking density for tRNA species but instead displaying high occupancy of eEF2 and partial OST occupancy were combined (13.6 %) and subjected to a second round of 3D classification with a mask on the translocon region. This resulted in a class with high OST-occupancy (6.8 % of all ribosomal particles) which was selected and refined to high resolution (Reconstruction B1). The final reconstructions for programmed and non-programmed OST-containing RTCs (A1 and B1) were further refined with a mask on the membrane components to enhance the signal-to-noise in this region (Reconstructions A2 and B2). The translocon region of A2 was used for model building and structural analysis of Sec61 and OST. It was combined with the ribosome region of A1 to evaluate the OST-LSU interface. A composite map of reconstructions B1 (ribosome) and B2 (translocon region) was used for rigid body fitting of the final models and comparison of conformational states.

technique revealed that the weaker signal in the membrane components was largely due to flexibility of the ribosome with respect to the membrane. Concomitantly, the masked refinement approach led to a significant improvement of the local

resolution for the translocon components Sec61 and OST (see figure 3.9).

For the transmembrane region of the programmed population local resolution now ranged between 3.5 Å in the Sec61 region to 4.5 Å for the most distal OST TMH. The local resolution in the non-programmed OST containing class was slightly worse in this region and varied from 4.2 Å (Sec61) to 5.5 Å (OST peripheral helices) in the membrane. Besides, a drastic resolution drop for the luminal OST segment was observed. As a consequence, certain regions in the luminal OST domain could not be resolved beyond ≥ 7.5 Å, limiting map quality to rigid, alpha helical features. The locally refined electron density for the translocon region allowed to assign the

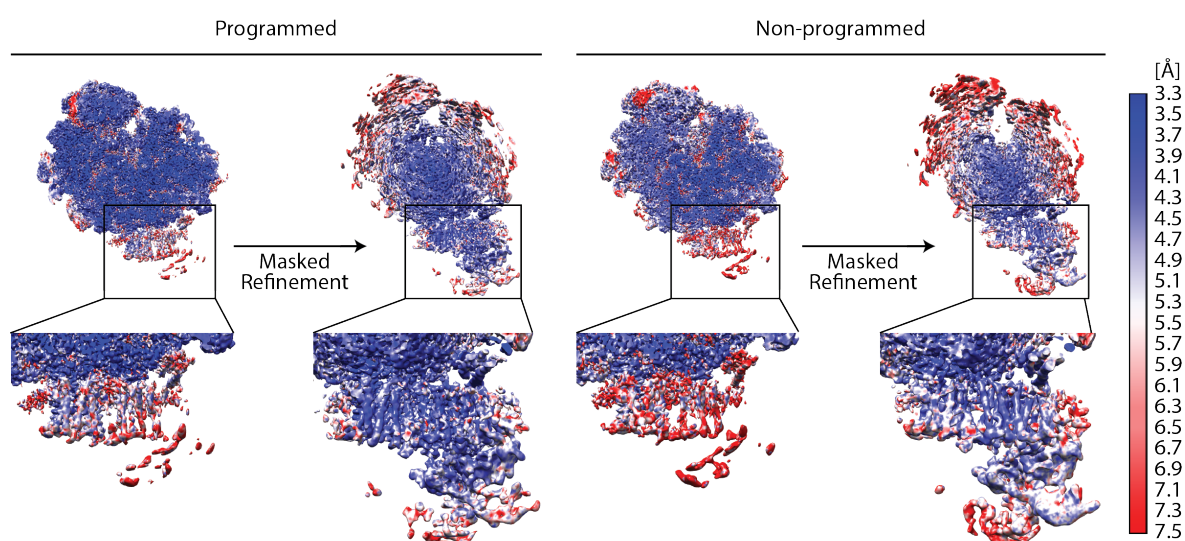


Figure 3.9 – Local resolution distribution. Maps of programmed and non-programmed particles before (left) and after (right) refinement on the ligand region colored according to local resolution. Local refinement significantly improved local resolution in the translocon region.

TMS of all OST components and to build a homology model of the catalytic subunit (described in detail in section 3.6). It was combined in a composite map with the high resolution electron density of the ribosome prior to masked refinement in order to visualize the general architecture of the OP96 biogenesis intermediate (see figure 3.10).

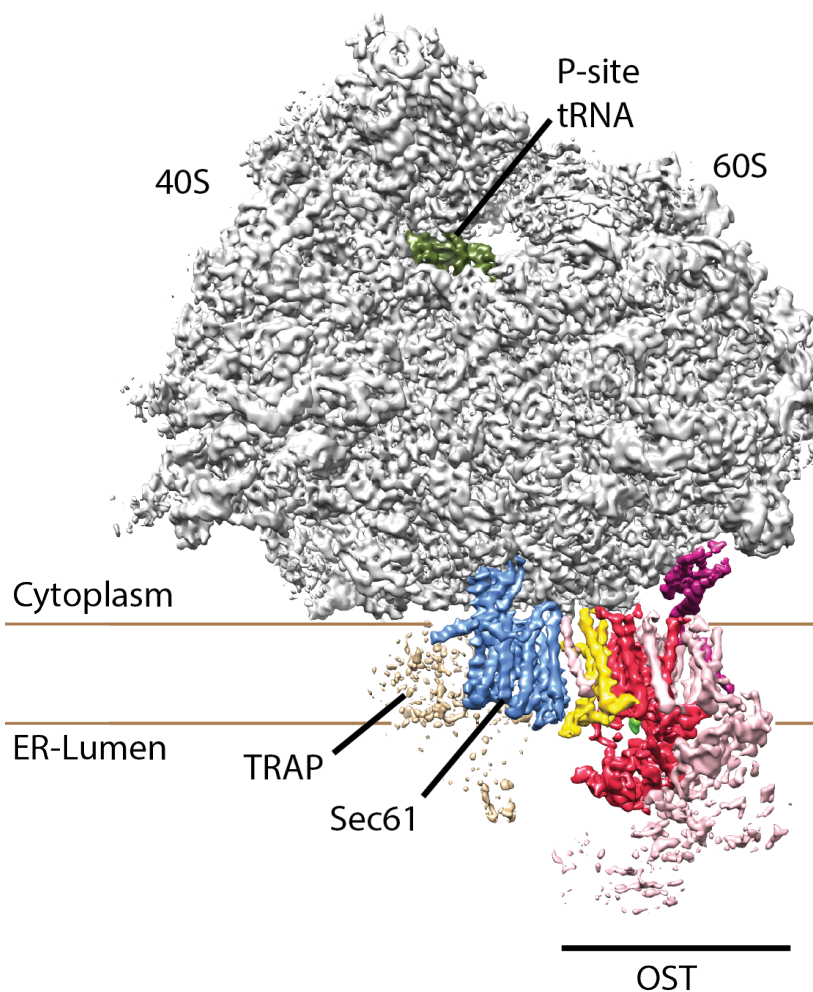


Figure 3.10 – Cryo-EM structure of the programmed solubilized RTC. Ribosome and p-site tRNA are shown before focused refinement, low-pass filtered to 4 Å; the membrane region including Sec61, TRAP, and OST is shown after. Horizontal lines indicate membrane boundaries prior to solubilization.

3.5 Conformation of Sec61

Analysis of the density corresponding to Sec61 confirmed our unexpected findings from the preliminary cryo-EM reconstruction. Despite being in complex with a ribosome translating a membrane insertion substrate, the PCC was very similar to the primed state (Voorhees et al., 2014) (see figure 3.11).

The lateral gate was in a closed position and the plug largely obstructed the central pore of the channel. In addition, no extra density corresponding to an inserting substrate TMH was visible in proximity of the lateral gate. We therefore decided to focus our analysis on additional density components which were present in the OST region.

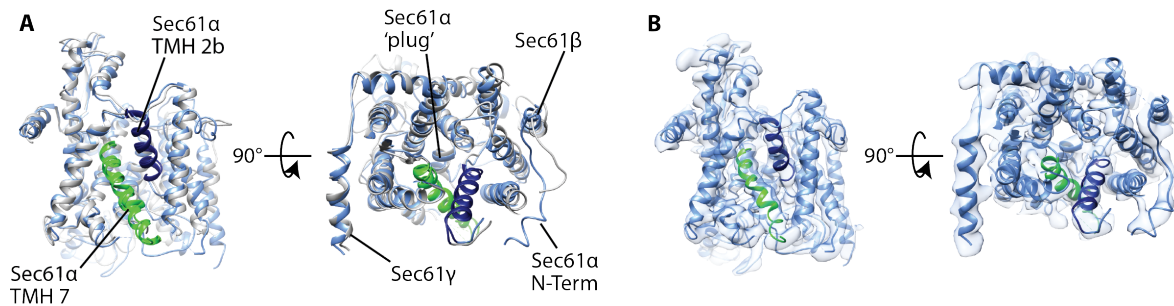


Figure 3.11 – Atomic model for Sec61. A) Overlay of the Sec61 model for the primed state (light grey, pdb 3j7q, Voorhees et al., 2014) with the adjusted model on basis of the high-resolution cryo-EM structure of the OP96 biogenesis intermediate (light blue); lateral gate TMHs are colored in dark blue (TMH 2) and green (TMH 7). Views perpendicular to the membrane plane (left) and from the cytosol (right) are shown B) Adjusted model of Sec61 as in A) in overlay with the corresponding cryo-EM density low-pass filtered to 4 Å according to local resolution.

3.6 Characterizing mammalian ribosome-bound OST

3.6.1 Evaluating OST-isoform specificity

We performed MS analysis of the OP96 cryo-EM sample in order to evaluate its protein composition prior to vitrification. The analysis identified a large number of ribosomal proteins and several factors involved in protein translocation and co-translational modification processes (See appendix A for a full list of identified proteins). Detected protein species include Sec61α and β, SRα and β, TRAM, TRAPγ and δ, signal peptidase subunits (Sec11, 25 kDa subunit) as well as components of the ER membrane protein complex (EMC1, 2, 3 5, 7 and 10; see Jonikas et al., 2009; Christianson et al., 2012). Finally, full sets of OST subunits for both complex isoforms were identified except for the OST4 component (see table 3.1).

Lack of OST4 could be explained by its small size, preventing it to be digested into peptide fragments suitable for MS detection. We were then wondering whether the MS analysis can provide information about the relative abundance of the two isoforms in our cryo-EM structure. Both OST isoforms are capable of co-translational N-glycosylation. However, biochemical data indicates that only STT3A-OST is associated with the ribosome (Shibatani et al., 2005; Ruiz-Canada et al., 2009). In line with this, the STT3B-OST specific subunits, STT3B and MAGT1 were detected at lower abundance than all other OST components. However, relative amounts of Sec61α were even lower according to the emPAI value (0.13), in contrast to the finding that cryo-EM analysis revealed a nearly complete occupancy of the PCC on ribosomal complexes (see figure 3.8). Therefore, apparent abundance according to MS

Table 3.1 – Abundance of OST subunits in the solubilized OP96 biogenesis intermediate. OST subunits in the OP96 cryo-EM sample by mass spectrometry, ranked according to their abundance by their emPAI values (exponentially modified protein abundance index, see Ishihama et al., 2005, Full range of observed values: 0.00-8.91).

Protein	Accession	emPAI
RPN1	gi 73984484	2.88
DAD1	gi 73962567	0.61
RPN2	gi 73991908	0.52
STT3A	gi 545496900	0.43
OST48	gi 164038	0.38
TMEM258	gi 345783280	0.38
KCP2	gi 226732647	0.22
DC2	gi 226739213	0.20
MAGT1	gi 345807587	0.16
STT3B	gi 73989883	0.15

analysis did not appear to be a suitable tool to evaluate which OST variant forms part of the structure.

Thus, in order to evaluate, whether the electron density which was observed in the high-resolution cryo-EM structure of the OP96 intermediate could be clearly assigned to one of the two paralogs we collaborated with the groups of Prof. Friedrich Förster (Utrecht University, Netherlands) and Prof. Reid Gilmore/Prof. Elisabet Mandon (University of Massachusetts, USA). Shiteshu Shrimal from the Mandon group isolated microsomal membranes from three different HEK-cell lines which were either wild-type (wt) or CRISPR-knockout mutants of STT3A or STT3B respectively (Shrimal et al., 2017). Subsequently Stefan Pfeffer from the Förster Lab performed CET in combination with subtomogram averaging on the wild-type HEK293, Δ STT3B and Δ STT3A microsomes. The data showed that wt and Δ STT3B had equal OST-occupancy for ER-associated ribosomes (see figure 3.12). In contrast Δ STT3 resulted in a loss of OST-containing complexes.

These results provided a structural proof that OST-containing RTCs harbor exclusively the STT3A-isoform of OST.

The CET data therefore also unambiguously identified the OST complex present in the high-resolution cryo-EM structure of the OP96 biogenesis intermediate as

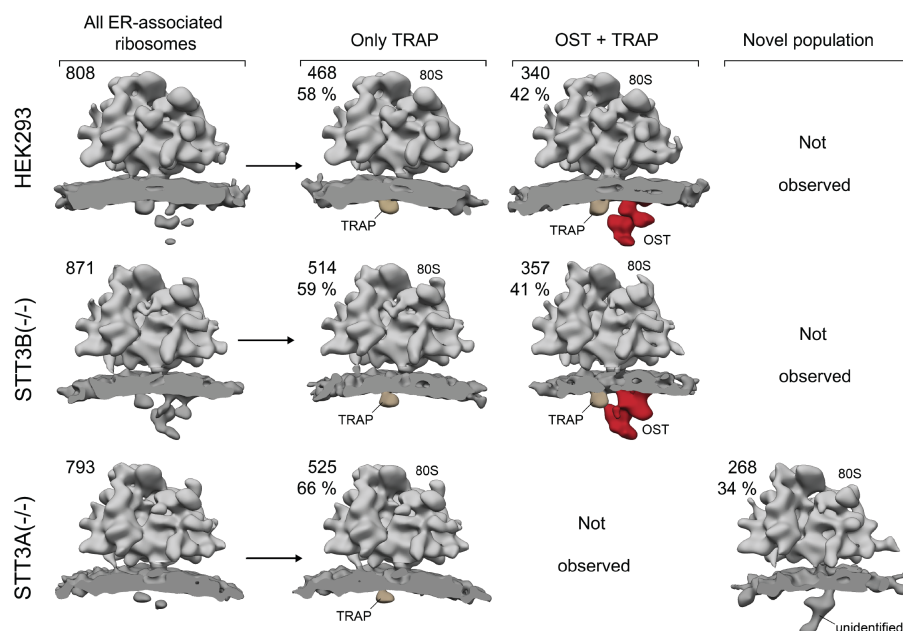


Figure 3.12 – CET of native RTCs. Ribosome-bound translocon populations observed by CET and subtomogram averaging for microsomes from wild-type HEK293 (top), Δ STT3B (center), and Δ STT3A (bottom) cell lines after in silico sorting. The absolute number and percentage of subtomograms contributing to each class are indicated. All densities were low-pass filtered to 30 Å (CET/subtomogram analysis and figure by Stefan Pfeffer).

STT3A-OST. This finding was not necessarily in contrast with detection of STT3B-OST specific components in the MS analysis. As mentioned previously, MS analysis also identified a range of additional other factors which were not accounted for by distinct density features (e.g. TRAP, EMC, SR). It is possible that these proteins were associated in the periphery, possibly in low amounts or rather transiently and therefore fell apart during the vitrification process. Being associated in proximity but not directly interacting with translocon-bound ribosomes is also in agreement with STT3B-OST's proofreading function for sites which have been skipped by STT3A.

3.6.2 Positioning of OST subunits

Having resolved potential ambiguity regarding the OST isoform present in the high resolution cryo-EM structure of the OP96 biogenesis intermediate we set out to identify the positions of OST subunits in the electron density region corresponding to the complex. The density envelope of the OST region was approximated from the known, medium-resolution CET structure of native, OST-bound RTCs (Pfeffer et al.,

2015). In our cryo-EM reconstruction of the OP96 biogenesis intermediate, 28 additional TMHs packed against Sec61, within the predicted OST-region. According to secondary structure predictions 31 TMHs would be expected for STT3A-OST indicating that at least one of the subunits was lost during preparation. We could unambiguously place a homology model of the catalytic STT3A subunit in the central region of the OST density (see figures 3.13 and 3.14). The model confirmed

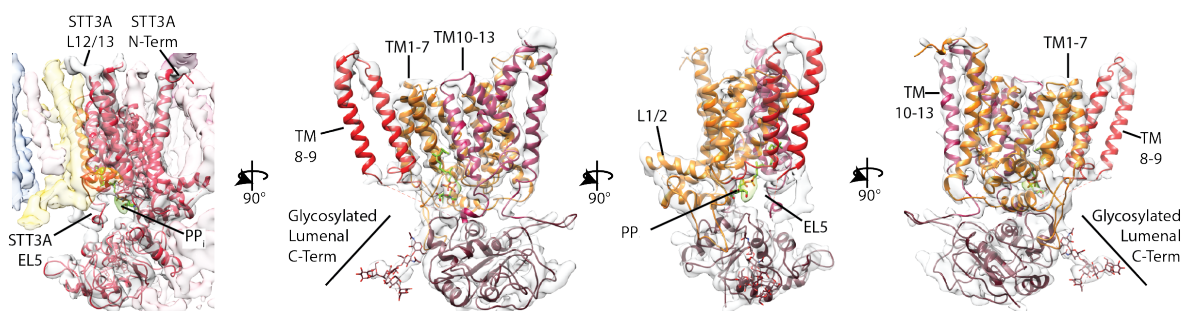


Figure 3.13 – Atomic model of mammalian STT3A. A model for canine STT3A was derived from a homology model based on existing crystal structures of the archaeal homolog and could be placed unambiguously in the electron density corresponding to OST.

the enzyme's 13 TMH topology and showed high similarity to the existing crystal structures of its homologs (see figure 3.14).

The most prominent rearrangements were observed in TMHs 8 and 9. In line with conformational flexibility in this region, TMH 9 was also the least resolved and no clear connectivity to the N-terminal part of EL5 was observed. Weak density for an α -helical segment was present in the region which likely represents the N-terminal end of EL5, but the residues involved could not be identified due to resolution restraints. In contrast, the C-terminal end of EL5 was relatively well structured starting from a conserved SVSE motif (TIXE in bacteria, with the glutamate constituting one of the residues of the active-site acidic cluster, see section 1.4.3). No density for the peptide substrate was observed whereas density corresponding to the pyrophosphate moiety of the LLO was present in the final reconstruction.

The remaining OST subunits assembled around the catalytic core and form clusters according to their affiliation with the three predicted subcomplexes (see figure 3.15).

A three-TMH bundle resolved at the distal side of the OST complex facing away from Sec61 and in close proximity to STT3A TMH1 corresponded to subcomplex I comprising RPN1 (1 TMH) and TMEM258 (2 TMHs). Subcomplex II comprises

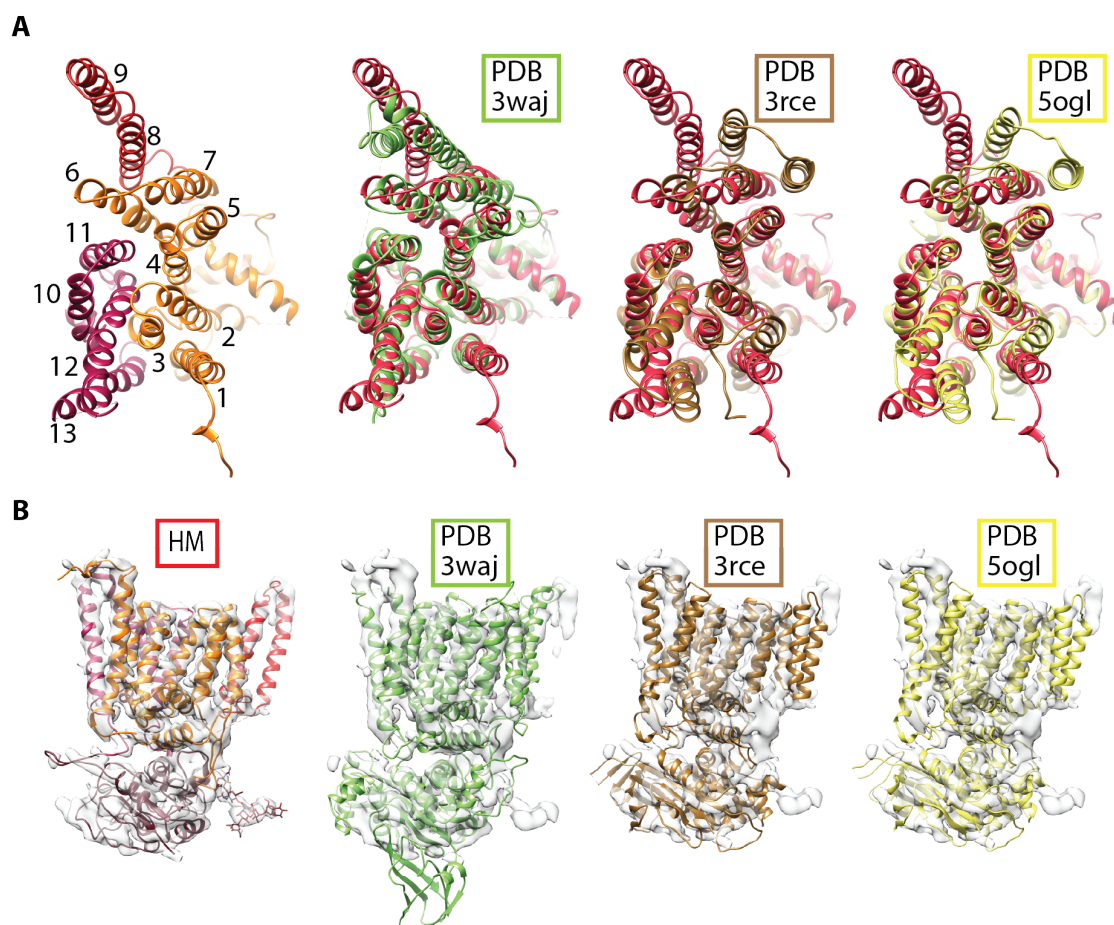


Figure 3.14 – Structural conservation of STT3-type proteins. A) Model for mammalian STT3A viewed from the cytosol (left, order of TMS is indicated by Arabic numbering: TMS1-7 orange, TMS8/9 light red, TMS10-13 dark red) and superposition with crystal structures of archaeal (pdb 3waj, Matsumoto et al., 2013) and bacterial (pdb 3rce, Lizak et al., 2011; pdb 5ogl, Napiórkowska et al., 2017) homologs. B) Overlay of the models from A) with the corresponding electron density in the high-resolution cryo-EM structure of the OP96 construct. Despite limited levels of sequence conservation a high degree of structural conservation can be observed.

STT3A (13 TMs), DC2 (3 TMHs), OST4 (1 TMH) and KCP2 (4 TMHs) and was located in immediate vicinity of the PCC. No extra density for the four TMHs of KCP2 was observed. The subunit has a known tendency to dissociate during solubilization (Shrimal et al., 2017) and could in addition have been lost during sample vitrification. Subcomplex III is composed of RPN2 (3 TMHs), DAD1 (3 TMHs) and OST48 (1 TMH) and was covered by the seven-TMH bundle on the distal side of STT3A. The large soluble extensions of subcomplexes I (RPN1) and III (OST48 and RPN2) enclosed the C-terminal domain of STT3A from opposite sides and approached each other below the catalytic subunit in the ER lumen. However the local resolution did not allow to build structural models for these domains.

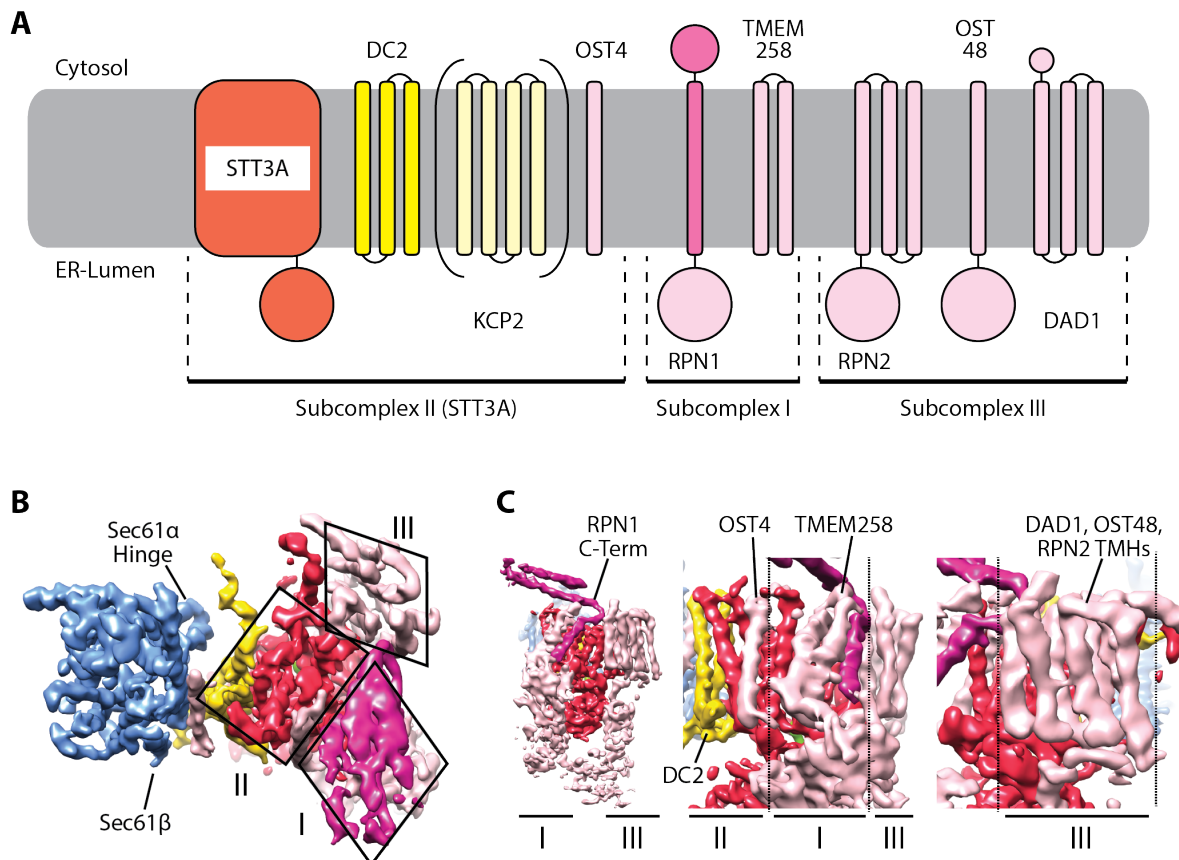


Figure 3.15 – Spatial arrangement of mammalian OST subunits. A) Schematic depiction of STT3A-OST subunits and predicted affiliation with subcomplexes I-III; KCP2 in parenthesis since no corresponding density could be identified in the OP96 cryo-EM reconstruction B) Cytosolic view of the Sec61-OST region in the OP96 cryo-EM map revealed a spatial organisation of STT3A-OST into the three biochemically predicted subcomplexes. The paralog specific components cluster in proximity to Sec61 whereas other subunits occupy more peripheral regions C) Arrangement of STT3A-OST subcomplexes when viewed parallel to the membrane plane.

3.7 Interactions of OST with the ribosome and Sec61

OST was integrated into the RTC via interactions through its RPN1 and DC2 subunits. The composite map as shown in figure 3.10 was used in order to draw conclusions on the interactions between the translation, translocation and glycosylation machineries.

3.7.1 The OST subunit DC2 at the Sec61 interface

The isoform specific DC2 subunit formed the interface to the Sec61 complex (see figure 3.16). The resolution in this area was sufficient to build an atomic model *de novo* for the three TMHs of DC2 based on excellent agreement between features resolved

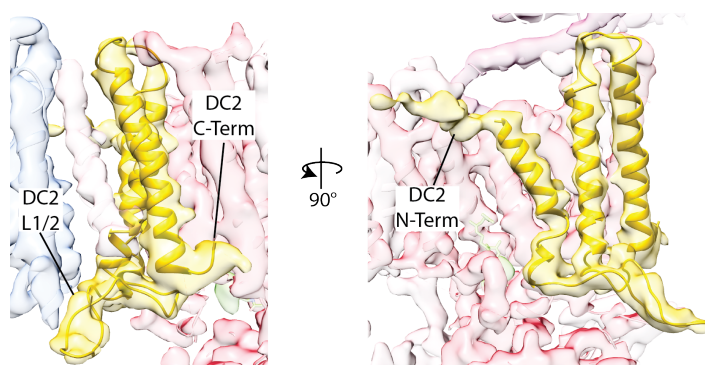


Figure 3.16 – The Sec61-DC2 interface. Magnified view of the Sec61-OST interface with a fitted model for the DC2 TMHs.

in our map and the predicted length and connectivity of DC2 TMHs. The amphipathic DC2 N-terminus projected toward Sec61 on the micelle surface, whereas the luminal loop of DC2 interacted with the C-termini of Sec61 β and Sec61 γ . DC2 contacted STT3A via its luminal C-terminus (to STT3A TMH13), the cytosolic L2/3 loop (to STT3A L12/13), and TM2 (close proximity to STT3A L9/10, also referred to as EL5). In the programmed population we observed an additional helical segment intercalated between Sec61 and DC2 which will be discussed in more detail in section 3.8.

3.7.2 The RPN1-Ribosome interface

The C-terminal region of RPN1 formed a four-helix bundle on the cytosolic face of the membrane and was intercalated between OST and the 60S ribosomal subunit. This metazoan-specific extension reached towards the linker between rRNA helix

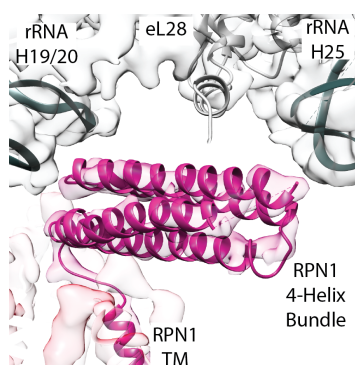


Figure 3.17 – The ribosome-RPN1 interface. Magnified view of the cytosolic RPN1 four-helix bundle binding to the ribosome.

H19 and H20, rRNA expansion segment ES7a (H25), and the tail of ribosomal protein eL28 which was unstructured in previous LSU models (Brown et al., 2015). Precise contact regions could not be defined since the LSU and the RPN1 bundle form part of two different density segments of the composite map.

3.7.3 Conformational dynamics in different functional states

The translocon is capable of dealing with a very broad range of substrates with different requirements regarding NC modifications, membrane insertion and topology determination. To gain insights into different functional states of the assembly we compared the conformation of Sec61 and OST in our model of the programmed OP96 biogenesis intermediate (maps A1 and A2) with their arrangement in the non-programmed counterpart (see B1 and B2 in figure 3.8) and in the native translocon (CET by Stefan Pfeffer, MPI of Biochemistry, Martinsried, improved processing routine for the data published in Pfeffer et al., 2015) (see figure 3.18).

The general translocon architecture in the non-programmed reconstruction was

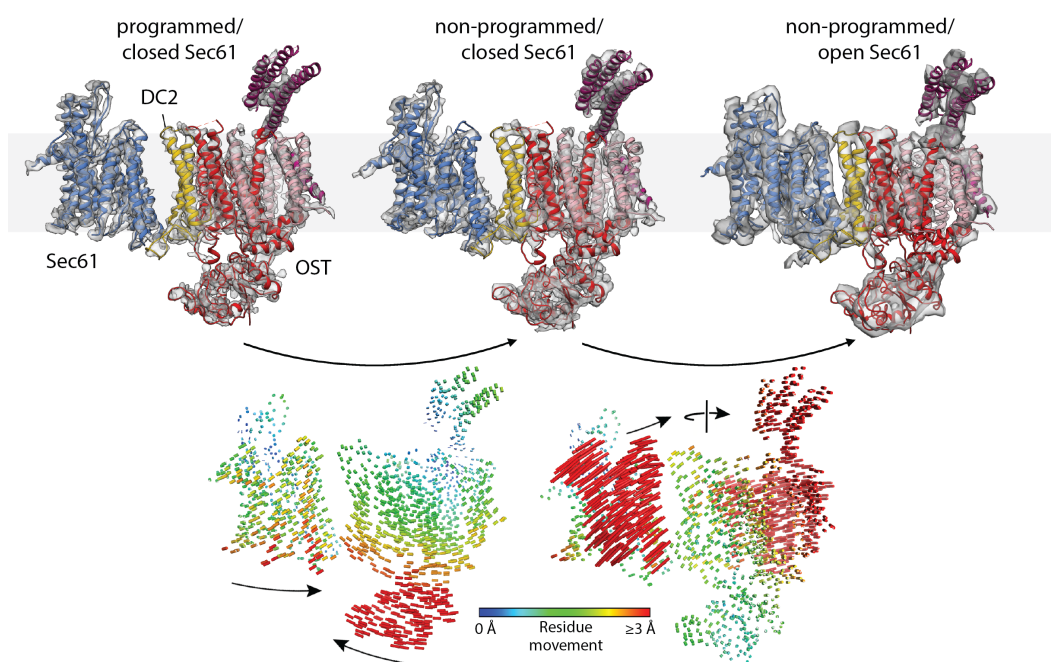


Figure 3.18 – Conformational states of different Sec61-OST complexes. Top: Rigid body fitted models for Sec61 and OST in the cryo-EM densities of the OP96 biogenesis intermediate (left, programmed; center, non-programmed) with laterally closed Sec61 and the native translocon from CET/subtomogram averaging with opened Sec61 (right). Bottom: Movement trajectories of α -carbon atoms connecting the observed conformational states with color-coded length (Figure by Stefan Pfeffer with small adaptations).

very similar to the P-site peptidyl-tRNA-containing complex, and models for laterally closed Sec61 and OST fitted well as separate rigid bodies (see figure 3.18, center). Comparison with the model of the programmed RTC revealed a tilt between Sec61 and OST, with the more luminal regions approaching each other. In this movement, the cytosolic loops of Sec61 and the cytosolic RPN1 domain served as hinge points on the LSU.

In the native RTC with laterally opened Sec61, the Sec61 α N-terminal domain and Sec61 β approached DC2. This induced a repositioning of the entire OST complex to accommodate the conformational change of the PCC.

Although the relative arrangement of OST and Sec61 differed substantially between the three observed conformational states, DC2 always mediated stable interactions with Sec61.

3.8 Tentative positioning of the nascent insertion-glycosylation substrate

Notably, after positioning of all OST subunits one additional helical density segment was observed intercalated between Sec61 and DC2 in the P-site peptidyl-tRNA bound OP96 sample (see figure 3.19, undisplayed for clarity in figure 3.18). The density was absent from the non-programmed population and its position was partially occupied as a result of the conformational rearrangements described in section 3.7.3.

This observation led to the idea that the density might represent TMH1 of the opsin *in vitro* translation/N-glycosylation substrate. To test this hypothesis, residues 35-59 of bovine opsin were tentatively modelled into the density (see figure 3.20). The model showed good agreement with electron density features for some bulky side-chains (e.g. W35, F56). Furthermore, there were no electron densities which immediately contradict the molecular model. However side-chain density could not be observed for all aa.

Since we isolated a co-translational biogenesis intermediate, the NC should still be connected to the p-site tRNA. Therefore, we evaluated the connectivity of the additional density segment in map A2 to the density observed in the ribosomal exit tunnel of the electron density map A1 obtained before masked refinement (see figure 3.21).

The NC displayed a flexible behavior, especially towards the lower part of the exit tunnel. Therefore, initially no connection between the two densities was observed.

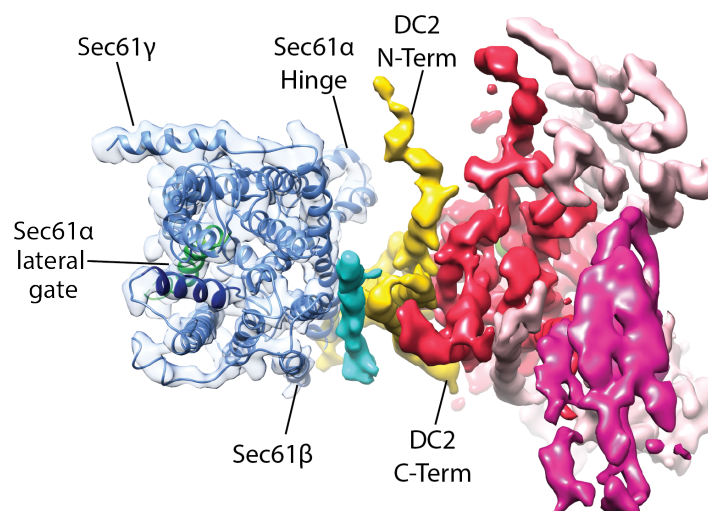


Figure 3.19 – Additional density at the Sec61-OST interface. An additional, unassigned helical density segment (cyan) is intercalated between Sec61 and OST in the programmed population of the OP96 cryo-EM sample on the opposite side of the lateral gate, view from the cytosol.

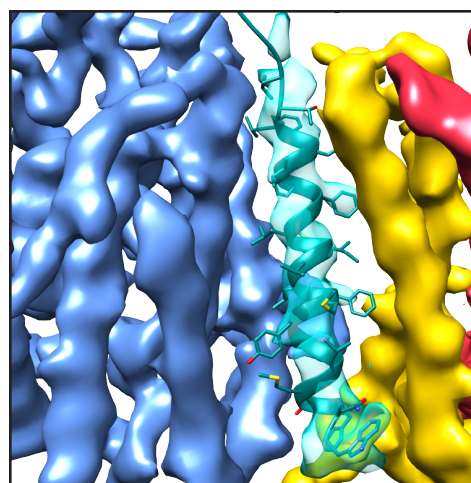


Figure 3.20 – Putative placement of the NC TMH. Close-up view on the unassigned density at the Sec61-OST interface, filtered to 3.5 Å and superposition with a model for opsin TMH1; sidechains correspond to opsin W35-L59.

However, upon decreasing the contour level for the isolated NC density in the ribosomal tunnel it could be traced to the tip of Sec61 TMH10. Additional density protruded from this connection point to the putatively identified substrate TMH. The above-mentioned results provided strong indication that the additional density segment could in fact be attributed to the first TMH of the *in vitro* insertion/N-glycosylation substrate opsin. However, for a PCC substrate, the density was positioned at a highly unexpected position. All known structures of insertion intermediates, both for the eukaryotic and the bacterial system display additional

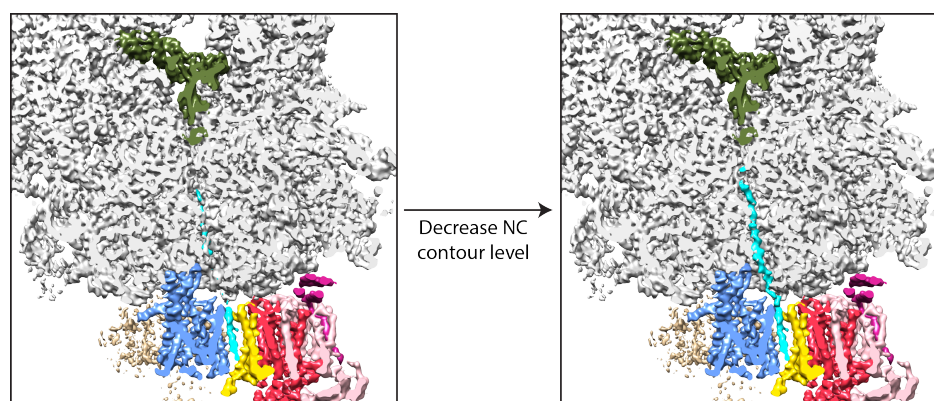


Figure 3.21 – OP96 NC connectivity. The electron density corresponding to the NC (cyan) connects to the putative opsin helix at low contour levels

density for the NC substrate intercalated at the Sec61 α lateral gate (Bischoff et al., 2014; Park et al., 2014; Gogala et al., 2014; Voorhees and Hegde, 2016). In contrast, the density segment in this study was observed almost on the opposite side of the channel, in the vicinity of Sec61 β . Remarkably, this position would result in a clash with Sec61 α TMH1 upon opening of the lateral gate. The OP96 construct contained only a fairly short linker between the C-terminal end of opsin TMH 1 (V64) and the PTC (32 aa). In order to rule out that the unusual positioning of the putative substrate helix was simply a result of distance restraints imposed by NC length, we decided to perform extended cryo-EM analysis on the OP109 construct. The construct has additional 13 aa and should therefore enable to relax potential strains affecting NC positioning. Hence, a second, larger cryo-EM dataset (dataset V, 12656 micrographs) was recorded for the OP109 construct and was processed by Robert Buschauer (Gene Center, LMU Munich) in analogy to the large OP96 dataset. However, NC elongation did not result in any structural changes and thus the data is not discussed in more detail. Therefore, it was concluded that the unexpected position of the tentatively identified opsin TMH is not an artefact caused by NC length.

In summary, we were able to derive a protocol for the isolation of *bona fide* mammalian membrane insertion intermediates of different nascent chain length, suitable for structural analysis. We solved the structure of a very early bovine opsin biogenesis intermediate representing the first high-resolution structure of a mammalian OST-bound RTC. We unambiguously identified the STT3A-isoform of OST and positioned the TMDs of all corresponding OST subunits, except for KCP2 which did not form part of the reconstruction. The arrangement revealed clustering of subunits according to three subcomplexes which had previously been proposed for *S. cerevisiae*

OST. Moreover, we derived a homology model for the catalytic STT3A subunit, *de novo* built an atomic model for the DC2 TMDs forming the interface with the Sec61 complex, and gained insights into OST ribosome binding via the RPN1 C-terminus. Finally, we surprisingly observed Sec61 in a near-primed state with a closed lateral gate and no additional insertion substrate density in the lateral gate region. Instead, we putatively identified the TMH of the opsin insertion/N-glycosylation substrate at an unexpected position between the PCC and OST.

Chapter 4

Discussion

A major goal of this work was to investigate the membrane insertion process of the chosen model substrate bovine opsin. We successfully established a protocol which enables the isolation of opsin biogenesis intermediates at different stages during the insertion process. In contrast to other structural studies of eukaryotic, solubilized RTCs, the purification routine is suitable to enrich for the oligosaccharyltransferase complex (OST), one of the translocon components which previously could only be observed in substoichiometric amounts (Gogala et al., 2014; Voorhees et al., 2014; Voorhees and Hegde, 2016). This provided the unique opportunity to investigate the intimate coupling of membrane protein insertion and asparagine-linked glycosylation (N-glycosylation) in a single structural study on functional opsin biogenesis intermediates.

4.1 A model for OST isoform specificity

N-glycosylation is a highly abundant, universally conserved modification of newly synthesized proteins, with great functional importance for protein folding, intracellular trafficking and stability. The modifying enzyme OST is a single subunit enzyme in prokaryotes and archaea but has evolved into a multisubunit assembly with at least eight different proteins in eukaryotes. Furthermore, gene duplication of the gene coding for the catalytic STT3 subunit in metazoans has yielded two complex isoforms, STT3A-OST and STT3B-OST. The majority of subunits are present in both variants and are complemented with either STT3A or STT3B and one or two paralog-specific subunits. Current models suggest based on biochemical data that only STT3A-OST associates with translocon-bound ribosomes for co-translational N-glycosylation. However, in addition to performing post-translational modifications, STT3B-OST can also act co-translationally. As part of this study, CET data

by Stefan Pfeffer provided the first structural proof, that ribosome-translocon binding is an exclusive feature of STT3A-OST in mammalian cells. We used cryo-EM and single-particle analysis to solve the first high resolution structure of an OST-containing RTC and obtained detailed insights into the STT3A-OST interfaces with Sec61 and the ribosome.

The PCC interaction is mediated by the isoform specific DC2 subunit as recently suggested by biochemical data (Shrimal et al., 2017). DC2 extends towards the Sec61 β and Sec61 γ subunits. Comparison of different functional RTC states demonstrated that DC2 acts as a versatile module which provides robust integration of OST into the translocon complex, even under vastly differing conformational states of the translocon complex. This is an important prerequisite for successful N-glycosylation of a very broad substrate range. Not only do OST substrates have vastly different sequon environments regarding their primary amino acid sequence. In addition they potentially require transient association not only with Sec61 and OST but also other accessory factors for nascent chain processing (signal peptidase), topogenesis (e.g. TRAM, TRAP) or covalent modification (e.g. protein disulfide isomerase). With regard to OST, DC2 is positioned adjacent to the most C-terminal helices of the STT3A TM region (TMH 10-13). This posed the question how DC2 is able to discriminate between the STT3A and STT3B paralogs. STT3A and STT3B show high overall sequence conservation (60% sequence identity in humans). However, more thorough investigation of pairwise alignments of STT3A and STT3B paralogs from different organisms revealed that conservation is significantly reduced for the DC2 binding region (see table 4.1).

Therefore, STT3A potentially presents a binding surface clearly distinct from the STT3B surface and hence preferably integrates DC2 which modulates Sec61 interaction. On the contrary, the interface presented by STT3B is presumably more suitable for binding of TUSC3 or MagT1. In comparison to DC2, these factors contain an additional TMH as well as a luminal thioredoxin-domain. It cannot be excluded that these features might even have a repulsive effect on Sec61 binding, for example due to steric hindrance. However, due to the lack of high resolution structures of STT3B-OST this remains rather speculative.

Towards the cytosol, STT3A-OST interacts with the ribosome via the C-terminal four-helix bundle of RPN1 according to our cryo-EM reconstruction. This metazoan specific domain protrudes from the TM region in direct vicinity of the STT3A N-terminus. Notably, the soluble N-terminal region is significantly longer in mammalian STT3B (41 vs. 17 aa in *Homo sapiens*, *Canis lupus familiaris*, *Mus musculus*, 66

Table 4.1 – Conservation of primary sequences between STT3A and STT3B from various metazoan organisms. Sequence identity between STT3A and STT3B paralogs for the sequence ranges covering TMHs 1-9 (residues 7-346 of reference sequence NP_689926.1) or TMHs 10-13 (residues 347-501 of reference sequence NP_689926.1) are shown. TMHs 1-9 have a significantly higher degree of conservation (60% identity between paralogs) than the DC2-interacting TMHs 10-13 (39% identity between paralogs)(Pairwise alignment performed by Stefan Pfeffer).

Organism		Sequence IDs STT3A/B	Sequence Identity TMH 1-9	Sequence Identity TMH 10-13
<i>Mammalia</i>	<i>Homo sapiens</i>	NP_689926.1/ NP_849193.1	60.70%	39.34%
	<i>Mus musculus</i>	NP_032434.3/ NP_077184.2	60.70%	39.34%
	<i>Bos taurus</i>	NP_001039445.1/ NP_001091039.1	60.70%	39.34%
	<i>Sus scrofa</i>	NP_001230781.1/ XP_003132136.3	60.41%	39.34%
<i>Aves</i>	<i>Gallus gallus</i>	NP_001305920.1/ XP_015137059.1	59.82%	38.80%
<i>Reptilia</i>	<i>Chrysemys picta</i>	XP_005294744.1/ XP_005299331.1	59.82%	39.34%
<i>Amphibia</i>	<i>Xenopus laevis</i>	NP_001083986.1/ XP_018124788.1	60.11%	39.34%
<i>Osteichthyes</i>	<i>Astyanax mexicanus</i>	XP_007244780.1/ XP_022529444.1	58.94%	39.89%

vs. 17 aa in *Bos taurus*, 69 vs. 15 aa in *Sus scrofa*). Hence, it is possible, that the extended STT3B N-terminus disturbs the interaction between RPN1 and the ribosome. This hypothesis is in agreement with results which demonstrate that binding of a Fab fragment from an anti-RPN1-C-terminus antibody can abolish ribosome translocon binding (Yu et al., 1990).

Interestingly, insects lack the C-terminal cytoplasmic domain of RPN1 despite expressing two paralogs of the catalytic subunit (Kelleher and Gilmore, 2006). Therefore it can be speculated, that the ribosome/RPN1 interaction rather plays a complementary role in anchoring OST to the RTC while DC2 acts as the primary binding partner. Alternatively, a tight RTC-OST interaction might be dispensable in organisms lacking the cytoplasmic RPN1 domain.

Taken together, our findings lead to a model which could explain how metazoan

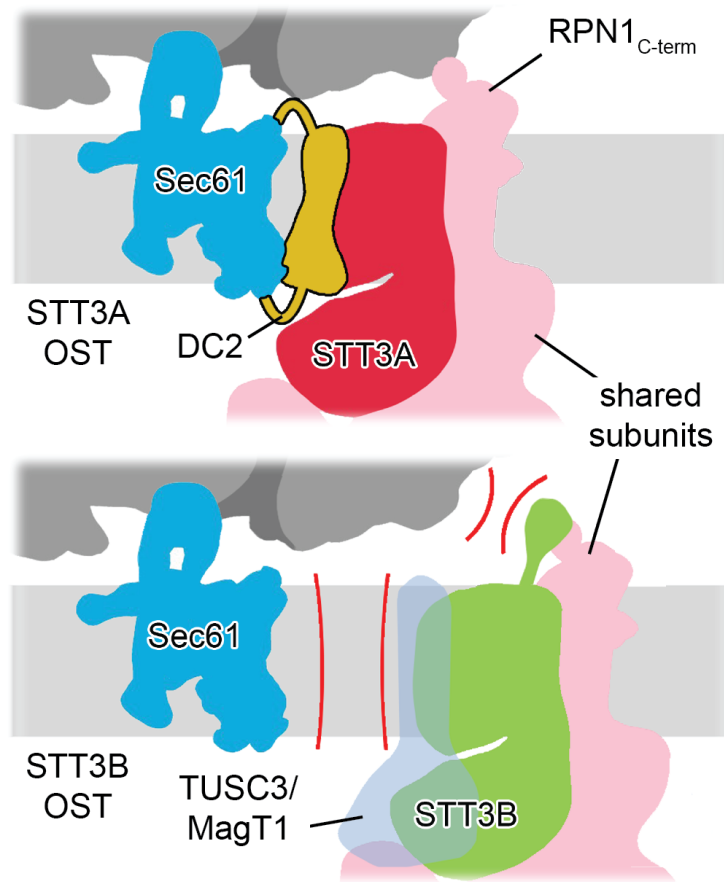


Figure 4.2 – Model for OST isoform specificity. Isoform specific features and components determine paralog specificity.(Figure adapted from Braunger et al., 2018).

RTCs selectively integrate the STT3A isoform of OST. It schematically summarized in figure 4.2 and represents a combination of positive determinants in STT3A-OST (DC2 integration) and negative determinants in STT3B-OST (N-terminal extension, L12/13).

4.2 Comparison of eukaryotic OST complexes

Prior to this study, the structural knowledge of the eukaryotic OST has been limited to moderate-resolution cryo-EM and CET structures (Li et al., 2008; Pfeffer et al., 2014; Pfeffer et al., 2015; Pfeffer et al., 2017) prohibiting a profound understanding of its subunit arrangement and interplay with the translation and translocation apparatus. The results presented in this dissertation provide the first high resolution model for the mammalian STT3A-OST in complex with Sec61 and the ribosome. In

addition, very recently, two other groups could successfully solve isolated cryo-EM structures of *S.cerevisiae* OST in the apo-state at 3.3 Å (Wild et al., 2018) and 3.5 Å resolution (Bai et al., 2018) respectively. They are very similar and provide molecular models for all yeast OST subunits except for the luminal domain and TMH1 of Ost3p and EL5 of Stt3p which were proposed to be mobile due to the absence of substrates.

In direct comparison the yeast and mammalian structures reveal a conserved architecture among eukaryotic OST complexes with an identical spatial division into subcomplexes in the TM region (see figure 4.3). Because of the high similarity of the *S. cerevisiae* structural models, only the one which is based on the slightly better resolved cryo-EM structure is shown (pdb 6ezn, Wild et al., 2018).

The fungal and metazoan complexes differ with regard to a few compositional el-

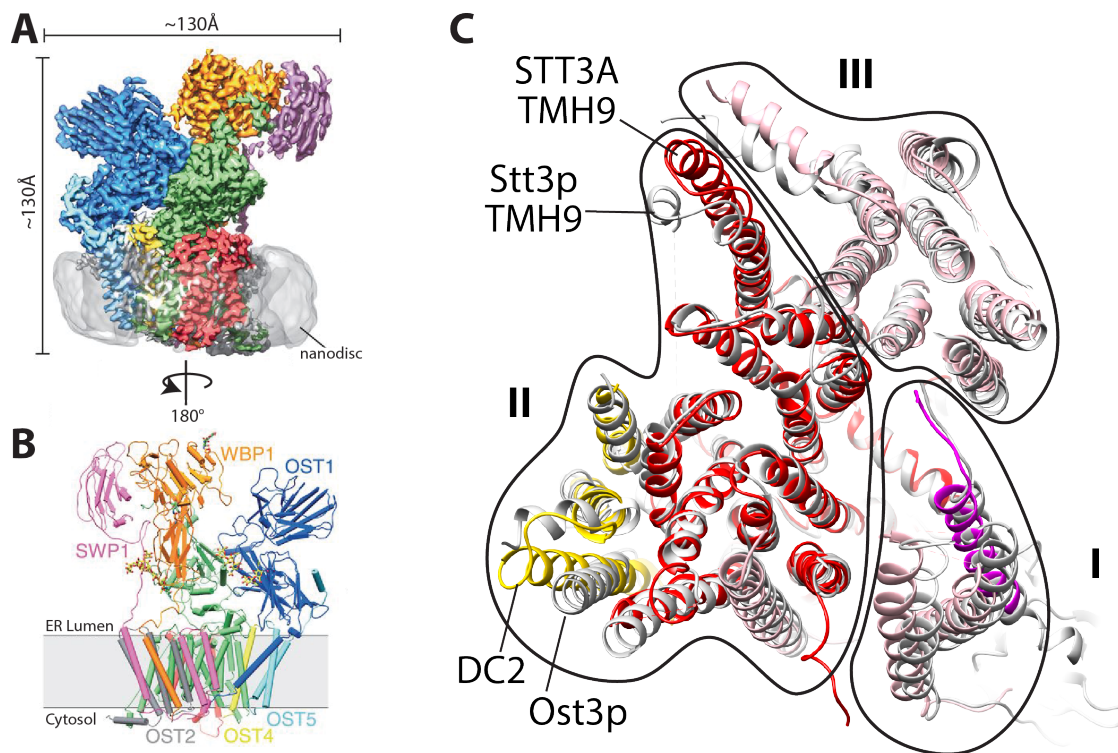


Figure 4.3 – Conserved architecture of eukaryotic OST complexes. A) Cryo-EM density for the OST complex from *S.cerevisiae*, reconstituted into nanodiscs (adapted from Wild et al., 2018), view in the membrane plane. B) Atomic model for the OST complex from *S.cerevisiae* based on the density shown in A), view rotated by 180° around the y-axis. C) Comparison of structural models for mammalian (pdb 6fti; DC2 in yellow, STT3A in red, RPN1 in magenta, other subunits in pale pink; Braunger et al., 2018) and yeast OST (pdb 6ezn; shown in white; Wild et al., 2018) demonstrates a conserved subunit organization in the membrane plane (view from the cytosol).

ements. DC2 engages the mammalian STT3A complex at the same position that is

occupied by the yeast Ost3p (TMHs 2-4), in agreement with the fact that yeast OST is more closely related to the mammalian STT3B isoform. Besides, the Ost1p segment which would correspond to the cytosolic four-helix bundle of RPN1 is missing in *S. cerevisiae*.

Interestingly, all three studies (Braunger et al., 2018; Bai et al., 2018; Wild et al., 2018) visualized a stably structured, highly conserved N-glycan linked to the catalytic subunit (to Asn539 in Stt3p/Asn548 in STT3A). The glycan forms interactions with the lumenal domains of subcomplex III (Wbp1p/Swp1p or OST48/RPN2 respectively) thereby bridging it with the STT3 subunit. The slightly better local resolution of the *S. cerevisiae* structures in addition revealed a number of ordered phospholipids forming the intra-membrane-interfaces between the three subcomplexes.

Within the STT3 subunit, the biggest difference between the yeast and mammalian models is the position of TMH9, which adopts a more outward oriented conformation in the mammalian STT3A structure (see figure 4.3). Together with STT3 TMHs 6 and 11 it forms a cavity surrounding the proposed dolichol binding site, hence different degrees of opening might reflect different states in the LLO binding cycle. As previously mentioned, the yeast structures represent the apo-state of OST without any bound substrates. The functional state of our mammalian STT3A-OST is less well defined. With regard to the peptide substrate, we do not observe any density in the sequon binding site. This is consistent with the fact that our isolated biogenesis intermediate reflects a post-glycosylation state, with both sequons in the opsin NC successfully modified according to WB analysis. However, it has to be considered that according to current models STT3A-OST employs a continuous N- to C-terminal scanning mode in order to sample nascent peptides for N-glycosylation sequons (Shrimal and Gilmore, 2013; Cherepanova et al., 2016). Thus, it would be expected that OST keeps interacting with the opsin NC even post-catalysis in order to probe for additional sites although the mode and site of these interactions remains to be characterized. A possible indication for ongoing nascent chain interactions the cryo-EM structure of the mammalian OP96 biogenesis intermediate is the fairly well structured conformation of the C-terminal half of EL5, starting from the conserved SVSE-motif. Though without any substrate density this idea remains rather speculative. The situation with regard to the LLO-binding state is similarly complex. We observe clear extra density located in the pyrophosphate binding pocket, but no densities for either the isoprenoid-carrier or the pre-assembled glycan. Hence, we cannot distinguish, whether the pyrophosphate density corresponds to the post-transfer PP_i -dolichol species prior to product release or whether the enzyme has

already recruited a new, fully assembled LLO-precursor, priming it for the next catalytic cycle. N-EL5 appears rather flexible and thus is not suitable to distinguish these two possibilities. In summary, the yeast and mammalian structures definitely represent different functional states and it seems plausible that these can be correlated to the position of STT3 TMH 9. However, the poorly defined state of the mammalian intermediate prevented a more detailed mechanistic analysis.

Remarkably, in contrast to both, the mammalian (Braunger et al., 2018) and the fungal models (Bai et al., 2018; Wild et al., 2018), STT3 TMHs 8 and 9 are shifted towards STT3 TMHs 5 and 7 in the bacterial PglB homolog (see figure 4.4) (Lizak et al., 2011; Napiórkowska et al., 2017).

Bai and colleagues speculate that the relocation of TMHs 8 and 9 in *S. cerevisiae* is a

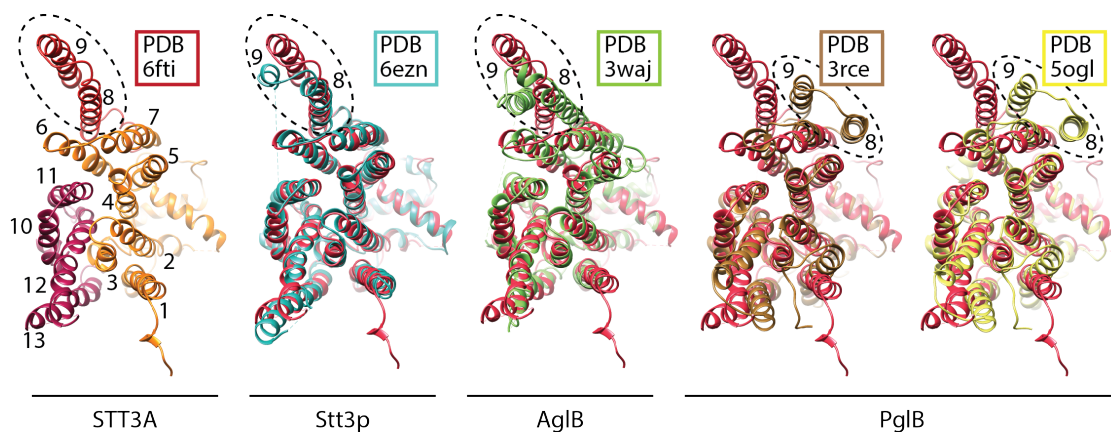


Figure 4.4 – Conformation of TMHs in STT3-type proteins. Comparison between structural models of metazoan (pdb 6fti, Braunger et al., 2018), fungal (pdb 6ezn, Wild et al., 2018), archaeal (pdb 3waj, Matsumoto et al., 2013) and bacterial (pdb 3rce/5ogl, Lizak et al., 2011; Napiórkowska et al., 2017) STT3-type enzymes reveals a highly conserved conformation except for TMHs 8 and 9 (dashed circle) which are drastically repositioned in the bacterial structures.

consequence of an altered LLO binding route due to the significantly increased size of the eukaryotic precursor (14-18 isoprene repeats and 14 sugars in yeast compared to 11 isoprene units and 7 sugars in most bacteria, see Jones et al., 2009). They formulate the hypothesis that the eukaryotic LLO dimensions prevent it from docking to the enzyme by diving under or threading in the EL5 loop as proposed for the bacterial PglB (Napiórkowska et al., 2017). Instead they suggest that it enters the binding cavity through a 10 Å gap between STT3 TMHs 8 and 9.

Can this model be supported by STT3 structures from the third domain of life? In crystal structures of the AglB homolog from *Archeoglobus fulgidus* (Matsumoto et al., 2013; Matsumoto et al., 2017) the enzyme's conformation more closely resembles

the eukaryotic variants (see figure 4.4). However, size-wise archaeal LLOs are more similar to the bacterial precursors with typically 10-12 isoprene units and short linear oligosaccharide chains (Jones et al., 2009). Therefore, the archaeal structures do not support the idea that TMH 8 and 9 repositioning is a direct consequence of LLO dimensions to allow for an alternative entry route. Nevertheless, they do not contradict the alternative LLO binding mode proposed by Bai and colleagues *per se*.

Due to limited local resolution we did not provide any structural models for the luminal domains of the non-catalytic subunits in mammalian STT3A-OST. In contrast, the better resolved yeast structures also allowed to build the luminal domains of Ost1p (RPN1), Wbp1p (OST48) and Swp1p (RPN2) and identified structural homologs for certain regions.

The luminal domain of Ost1p contained two structurally similar regions featuring a fold of stacked β -sheets and both groups identified resembling domains in aminopeptidases and leukotrien hydrolases. The two studies identified different structural homologs for the Wbp1p middle domain. While Bai *et al.* claimed similarity to amylase domain N, Wild and colleagues find homologies to protein folds in factors of the complement system of innate immunity. In addition, Bai and colleagues describe a structural similarity between the Swp1p N-terminal domain and a lipopolysaccharide- and sugar-binding co-receptor of a Toll-like receptor complex. The two groups draw different conclusions with respect to the functions of the extended luminal domains in non-catalytic OST subunits. Bai and colleagues conceive that the structural homologs which were identified in the Wbp1p and Swp1p N-terminal domains are sufficient to support their role in LLO recruitment. They do not propose any functional role for the Ost1p N-terminus. In contrast, Wild and colleagues rather favor a role for all three luminal domain in recruitment of additional factors such as chaperones or factors acting on nascent glycoproteins.

In summary, despite providing structural information, the *S. cerevisiae* model cannot provide conclusive data on the Ost1p, Wbp1p and Swp1p subunits' function in the N-glycosylation pathway. It cannot be excluded, that they have evolved to modulate N-glycoprotein biogenesis in an organism specific manner. This idea is supported by the fact that apart from few conserved patches, which were identified by Wild and colleagues and mediate subcomplex interactions, the luminal domains display low overall sequence conservation (see figure 4.5).

To the same end, the luminal domain of mammalian RPN2 harbors approximately 340 residues more than its yeast counterpart Swp1p, resulting in an additional lobe of electron density in the corresponding cryo-EM structure presented in this thesis

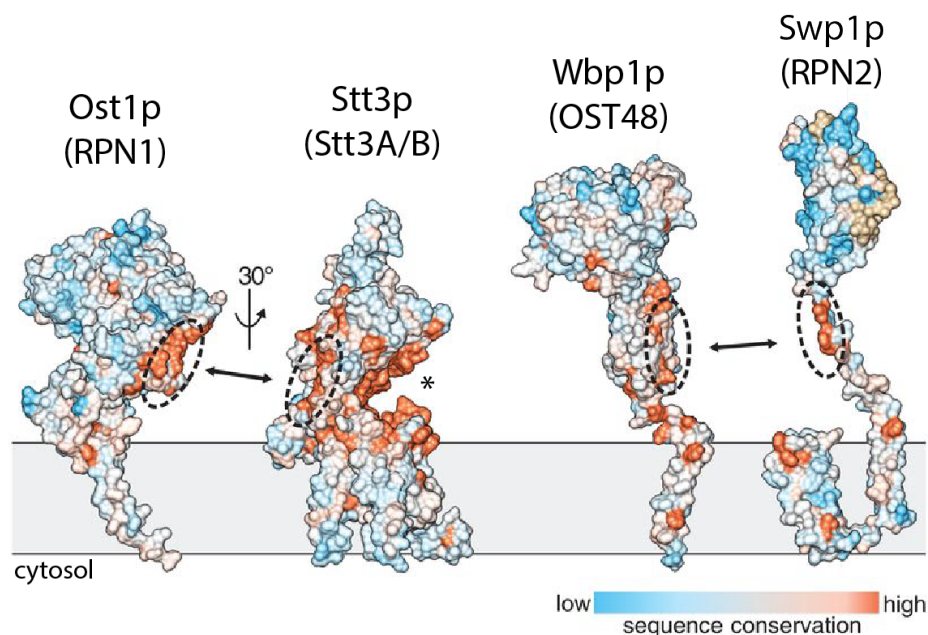


Figure 4.5 – Sequence conservation between different eukaryotic OST subunits. Sequence conservation among STT3-type and subcomplex III OST subunits of various eukaryotic organisms (*S. cerevisiae*, *P. pastoris*, *C. elegans*, *A. thaliana*, *D. melanogaster*, *D. rerio*, *M. musculus* and *H. sapiens*) reveals low overall conservation with few conserved patches at the substrate binding groove (asterisk) and the subunit interaction interfaces (dashed circles) (Figure adapted from Wild et al., 2018).

(see figure 4.6).

Therefore, the biological roles and mechanisms of function for the non-catalytic OST luminal domains remain ill-defined.

Finally, the question remains, whether *S. cerevisiae* OST is capable of engaging the RTC similar to the mammalian STT3A-OST despite higher similarity to the STT3B-type complex. The structures cannot provide an ultimate answer. Unfortunately the luminal domain of Ost3p as well as its most N-terminal TMH could not be resolved in the yeast structures, likely due to conformational flexibility. Bai *et al.* observe a weak density corresponding to the Ost3p thioredoxin domain in one of their classes. However, they use a different class without this extra density for docking into the CET structure of the mammalian OST-containing translocon which does not reveal sterical clashes in the membrane region. It is conceivable though, that an additional luminal domain in proximity to the position observed for Ost3p might lead to clashes with other translocon components such as the TRAP luminal domain or the Sec61 α hinge region which are in close proximity in the mammalian complex (see figure 4.6). Furthermore, Ost3p is lacking important structural features which characterize the Sec61-DC2 interaction in our mammalian model. The loop between

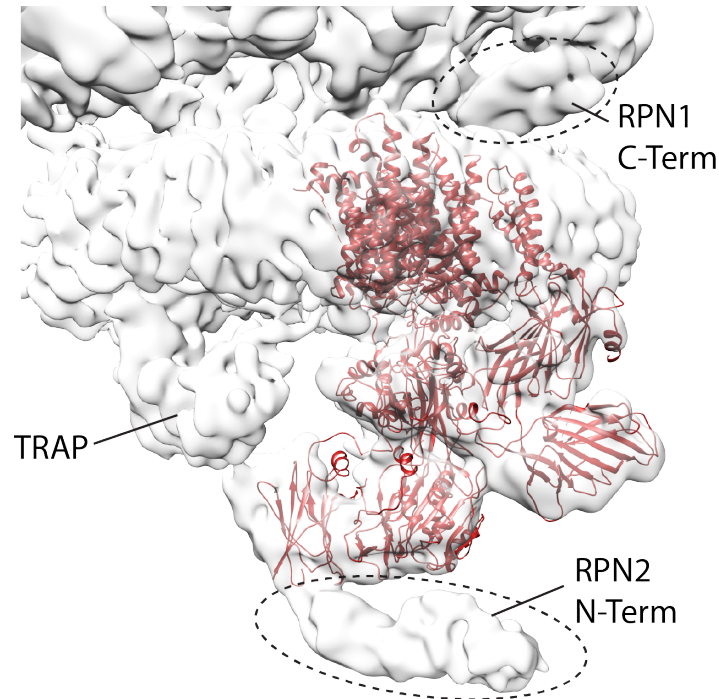


Figure 4.6 – Superposition of a model for fungal OST and the cryo-EM density of mammalian STT3A-OST. Model of the *S. cerevisiae* OST complex (red, pdb 6ezn, Wild et al., 2018) has been fitted as a rigid body into the cryo-EM density of the translocon region from the mammalian OST-containing RTC, filtered to 7 Å. Density corresponding to the mammalian STT3A-OST displays additional segments of density for the luminal RPN2 N-terminal and cytoplasmic RPN1 C-terminal domains.

Ost3p TMHs 2 and 3 is shorter and adopts a different fold than the corresponding L1/2 in DC2. Moreover, the DC2 N-terminal region is replaced by an additional TM span followed by the luminal domain. Taken together with the fact that *S. cerevisiae* does not contain the cytoplasmic ribosome-interacting domain of RPN1 in its homolog Ost1p (see figure 4.6), it is highly questionable whether the yeast OST complex is bound to RTCs *in vivo*. Therefore RTC association which allows immediate co-translational NC scanning appears to be a metazoan specific feature of STT3A-type-OST complexes.

4.3 Opsin membrane insertion takes unexpected ways

Three opsin biogenesis intermediates were isolated and analyzed by cryo-EM (OP96, OP109, OP204). Their anticipated architecture according to current knowledge on opsin topogenesis and membrane protein insertion in general is schematically illustrated in figure 4.7.

We were expecting 80S ribosomes which are programmed with a P-Site tRNA

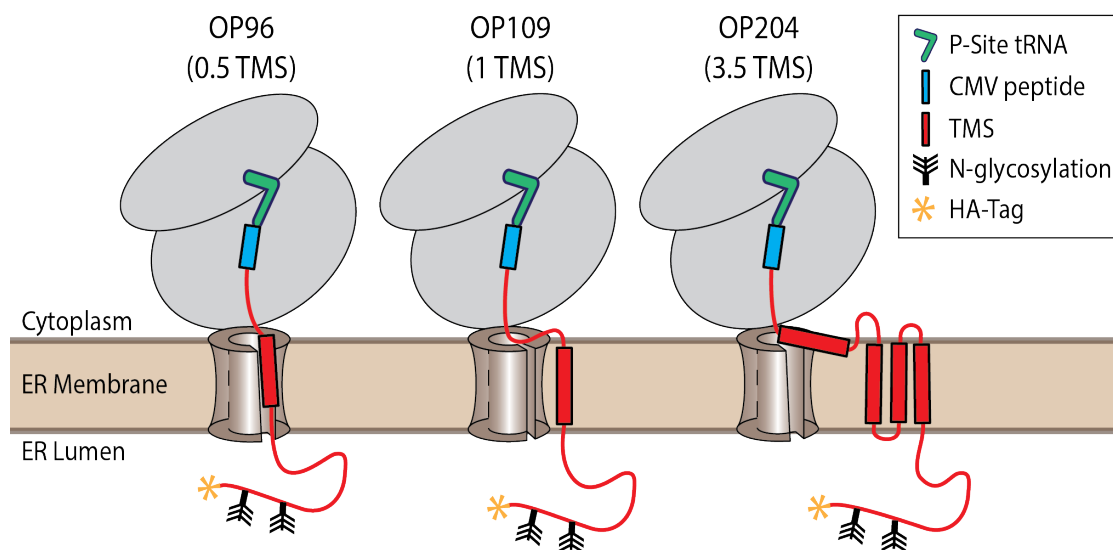


Figure 4.7 – Expected architecture of opsin biogenesis intermediates. Schematic depiction of the structural features expected in the cryo-EM structures of the OP96, OP109 and OP204 biogenesis intermediates. Indicated numbers of TMS correspond to the number of TMS which were expected in the membrane region at a given NC length.

bound to the corresponding nascent chain species. Previous studies have shown that the CMV stalling peptide which we included in our constructs for generation of stable, uniform RNC populations can form α -helical secondary structure in the upper part of the ribosomal tunnel (Matheisl et al., 2015). In the membrane region we predicted Sec61 in an inserting state with potentially varying degrees of lateral gate opening as well as helical extra densities for the inserting substrate TMHs intercalated into or in proximity of the gate. This assumption was based on available structures of inserting Sec61 (Gogala et al., 2014; Voorhees and Hegde, 2016) and the findings that opsin TMHs 1-3 form a PCC-proximal bundle prior to full release into the lipid bilayer (Ismail et al., 2006).

In the actual cryo-EM reconstructions a major class containing Sec61-bound RNCs with a P-site peptidyl tRNA and varying amounts of OST could be sorted out *in silico*. We did not detect a stable α -helical conformation of the CMV peptide in the ribosomal tunnel. In this context it is important to point out that the CMV peptide triggers stalling at a stop-codon. Concomitant with that, its structured conformation was visualized on a translation termination intermediate in complex with eukaryotic release factor 1 (eRF1, Matheisl et al., 2015). Apart from stable association with eRF1 the structure also revealed a role of the first nucleotide after the stop-codon triplet in formation of a stable termination complex. Hence it is

possible that in the study presented here, truncation of the mRNA immediately after the CMV STOP-codon, led to an altered PTC geometry which in turn influenced the conformation of the NC. Nevertheless, we visualized NC density in the entire ribosomal tunnel although local resolution did not allow to build an atomic NC model and unambiguously determine its conformation. Contrary to expectations, we did not observe any rod-like extra densities for the insertion substrate near the Sec61 lateral gate. Moreover, the PCC was in a non-inserting state, closely resembling the primed coformation for all reconstructions. Since all of the samples were successfully glycosylated according to the MW observed in WB detection of NCs it can be ruled out that Sec61 conformation is a result of unspecific membrane binding and/or improper PCC engagement.

Extended cryo-EM analysis of the OP96 biogenesis intermediate revealed an unassigned rod-like density intercalated between Sec61 and OST. We provisionally identified the additional helical segment as the TMH of the opsin *in vitro* insertion/N-glycosylation substrate. This positioning was highly unexpected since it is almost on the opposite side of the PCC than the lateral gate, in vicinity of Sec61 α TMH 1 and Sec61 β . However, it is consistent with existing crosslinking studies which described interactions between Sec61 β and the first opsin TMH until a NC length of 150 aa (Meacock et al., 2002; Ismail et al., 2006). Due to distance restraints these crosslinks are incompatible with positioning of TMH1 at the lateral gate.

Assuming our assignment is valid, there are two possibilities how the TMH has reached its position. Following the canonical in-out model of membrane insertion, the TMS would have to first engage the central cavity of Sec61, subsequently partition into the membrane through the lateral gate and finally relocate around the PCC. The sliding model provides an alternative explanation. Here, Sec61 might simply facilitate initial contact between the hydrophobic TMS and the lipid phase. Thermodynamically, it is then favorable for the TMH to insert directly into the membrane without sampling the hydrophilic central cavity of Sec61 beforehand. This spontaneous insertion could occur anywhere in proximity to the PCC, possibly depending on requirements of additional translocon components for topogenesis of the respective substrate.

Which explanation seems more likely for the OP96 biogenesis intermediate? The very short linker between the TMH C-terminus and the PCC (32 aa) greatly reduced the NCs mobility, arguing against the classical in-out model. In contrast, it is in favor of TMH sliding along the outer surface of Sec61 in direct vicinity of the OST in order to allow for successful glycosylation of the N15 which is only 20 aa N-terminal

of TMH1. Spontaneous insertion is furthermore supported by the fact that TMH1 is the most hydrophobic of all TMHs according to ΔG prediction (-2.44 kcal/mol) (Hessa et al., 2007). The conformation appears to be stable also with longer NC as indicated by the extended analysis of the OP109 intermediate. Therefore it must be taken into consideration that our cryo-EM structure represents an equilibrium state. Consequently, the observed positioning does not necessarily correspond to a kinetically relevant on-pathway intermediate *in vivo* but might rather reflect accomodation of the TMH at the thermodynamically most stable position.

Chapter 5

Future perspectives

The results presented in this dissertation provide the first high resolution structure of a mammalian co-translational membrane insertion/N-glycosylation intermediate. They yield important insights in particular with respect to the composition and architecture of translocon-bound OST complexes and form the base for a model which can explain the selective integration of STT3A-OST into the RTC of higher eukaryotes.

Nevertheless many open questions remain regarding the vital process of N-glycosylation. First, what are the mechanisms involved in sequon recognition for STT3A- and STT3B-OST? Available data suggest that the two complexes engage glycosylation substrates in fundamentally different ways (Cherepanova et al., 2016). As shown in this work, STT3A-OST is directly coupled to the protein synthesis and translocation machinery and is proposed to generically sample nascent chains by continuous scanning. In contrast, STT3B-OST is thought to selectively recruit its target peptides via a mechanism involving the thioredoxin domain of its paralog specific subunit in order to maximise N-glycan coverage.

Second, what is the role of the paralog-independent non-catalytic subunits, especially the large luminal domains of Ost1p/RPN1, Wbp1p/OST48 and Swp1p/RPN2? There are indications that they contribute to complex stability, substrate binding and selectivity. However, how these functions are distributed among subunits and how they accomplish these tasks remains to be clarified. One important step in this direction would be to complement the recent structural models of the *S. cerevisiae* proteins (Bai et al., 2018; Wild et al., 2018) with high resolution structures of their metazoan counterparts. To address this issue, recent developments in cryo-EM data processing software (Zivanov et al., 2018) such as CTF-refinement including various per-particle corrections and multi-body refinement might help to further increase the local resolution in the translocon region of the mammalian opsin biogenesis intermediates presented in this work.

Another objective of future research will be to characterize the substrate recruitment process in OST enzymes of the three kingdoms of life. On the one hand, all STT3-type proteins appear to share a common catalytic mechanism and have high structural similarity. On the other hand, certain properties especially concerning the size and structure of the LLO-precursor hint to potential differences in its delivery to the active site (Jones et al., 2009; Napiórkowska et al., 2017; Bai et al., 2018). The LLO-entry route and binding mode might also be clarified by further analysis regarding the different observed conformations of the EL5-loop. What is the functional importance of these conformational changes? Here, it is known that presence of LLO or peptide substrates induces stable conformations of either N-EL5 or C-EL5 respectively but the order of events and the exact triggers of conformational stabilization are unclear. Thus, it would be desirable to generate more defined eukaryotic N-glycosylation intermediates in different stages of the catalytic cycle for extended structural studies.

Aside from contributing valuable insights into the process of N-glycosylation, this thesis led to a number of unexpected findings with regard to the membrane insertion procedure of bovine opsin. How can the near-primed conformation of Sec61 be explained, which was observed for all three biogenesis intermediates presented in this work? Is the additional density at the Sec61-OST interface in the high resolution cryo-EM structures of OP96 and OP109 indeed representing the inserting substrate? If so, how did it reach this position? In order to address these questions, it is crucial to obtain proof of the tentative assignment and positioning of the insertion substrate. As mentioned above, a processing routine including the latest algorithmic developments might lead to improved definition of density features in the translocon region. Thus, it will potentially allow to confirm the identity of the tentatively assigned opsin TMS. An alternative approach would be targeted NC modifications such as additional bulky side chains or the incorporation of unnatural aa at defined positions to allow site specific crosslinking to proximal translocon components. In contrast to existing crosslinking studies the experiments could be coupled with mass spectrometry in order to reveal not only the identity of the NC interaction partner but also provide information on the interacting region. Furthermore, it will be particularly interesting to evaluate whether intermediates with a higher number of inserted TMS reside at the same position as provisionally shown for OP96 and OP109. To that end, extended analysis of insertion intermediates reflecting later stages during opsin biogenesis, such as the OP204 intermediate should be performed. Additionally, structural analysis of biogenesis intermediates from

other polytopic insertion substrates could help to reveal determinants for certain insertion trajectories.

Finally, the role of many accessory translocon factors remains in the dark. Until now, they evade structural analysis since they seem to dissociate during the sample preparation process. In our analysis TRAP, TRAM as well as the majority of EMC proteins were identified in the sample but could not be visualized at all or at sufficiently high resolution to draw any conclusions. Therefore, another future challenge lies in experimental stabilization of these transient assemblies to make them accessible for cryo-EM analysis.

In summary, this study once again demonstrates that the biogenesis of α -helical TMPs is a multi-dimensional problem with many remaining experimental challenges, especially, but not only considering polytopic substrates. It requires a dynamic network of PCC-associated factors, some of which might form only very transient interactions or perform their function in a highly substrate specific manner. Considering the tremendous biological significance of the inserted substrates, representing up to 30% of an organism's protein encoding genes (Von Heijne, 2011) membrane protein insertion and topogenesis continues to be a highly relevant field of study.

Appendix A

MS analysis of the solubilized OP96 biogenesis intermediate

A.1 Hits from the *canis lupus familiaris* database

Table A.1 – MS analysis of the solubilized OP96 sample - Hits from *canis lupus familiaris* (dog). Mass spectrometry analysis was performed by Thomas Fröhlich (Gene Center, LMU), hits are sorted according to their abundance as quantified by the emPAI value. All subunits from mammalian OST are highlighted in grey. Short names are given for all identified protein factors involved in translocation processes. The database from *canis lupus familiaris* was used for the identification of membrane components due to the use of dog PKRM in the translation reaction. Hits from soluble components arise from high similarity between *canis lupus familiaris* (PKRM) and *oryctolagus cuniculus* (RRL).

Accession	emPAI	Description	Short Name
gi 545488327	8.91	PREDICTED: 40S ribosomal protein S19 [Canis lupus familiaris]	
gi 73987191	8.19	PREDICTED: 40S ribosomal protein S28 [Canis lupus familiaris]	
gi 187609320	5.71	Chain u, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 55956952	4.70	ribosomal protein S18 [Canis lupus familiaris]	
gi 356582340	4.42	ubiquitin-40S ribosomal protein S27a [Canis lupus familiaris]	
gi 57113163	4.14	PREDICTED: 60S ribosomal protein L30 [Canis lupus familiaris]	
gi 187609309	4.13	Chain k, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 928169639	3.98	PREDICTED: 40S ribosomal protein S21 [Canis lupus familiaris]	
gi 545524789	3.80	PREDICTED: 40S ribosomal protein S3a [Canis lupus familiaris]	
gi 187609255	3.56	Chain a, Structure Of A Mammalian Ribosomal 40s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 187609317	3.23	Chain r, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 187609307	2.96	Chain i, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 187609304	2.91	Chain f, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 73984484	2.88	PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 [Canis lupus familiaris]	RPN1
gi 57101324	2.78	PREDICTED: 60S ribosomal protein L18a [Canis lupus familiaris]	
gi 187609264	2.71	Chain k, Structure Of A Mammalian Ribosomal 40s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 325530283	2.62	RecName: Full=40S ribosomal protein S3	
gi 73988614	2.51	PREDICTED: 40S ribosomal protein S13 [Canis lupus familiaris]	
gi 73974792	2.42	PREDICTED: 60S ribosomal protein L8 [Canis lupus familiaris]	
gi 984114	2.31	ribosome receptor [Canis lupus familiaris]	
gi 545497049	2.24	PREDICTED: heat shock cognate 71 kDa protein [Canis lupus familiaris]	
gi 187609269	2.22	Chain q, Structure Of A Mammalian Ribosomal 40s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 354725920	2.20	60S ribosomal protein L21 [Canis lupus familiaris]	
gi 345806081	2.17	PREDICTED: 78 kDa glucose-regulated protein [Canis lupus familiaris]	
gi 345780888	2.17	PREDICTED: 40S ribosomal protein S8 [Canis lupus familiaris]	
gi 44888811	2.15	RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglobin beta chain	
gi 73999396	2.11	PREDICTED: 40S ribosomal protein S20 [Canis lupus familiaris]	
gi 187609262	2.06	Chain i, Structure Of A Mammalian Ribosomal 40s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 73969959	1.93	PREDICTED: cytoskeleton-associated protein 4 [Canis lupus familiaris]	
gi 73945531	1.80	PREDICTED: 40S ribosomal protein S12 [Canis lupus familiaris]	
gi 345805550	1.69	PREDICTED: LOW QUALITY PROTEIN: leucine-rich repeat-containing protein 59 [Canis lupus familiaris]	
gi 545545654	1.63	PREDICTED: peptidyl-prolyl cis-trans isomerase FKBP11 [Canis lupus familiaris]	
gi 356582337	1.63	60S ribosomal protein L7 [Canis lupus familiaris]	
gi 356991214	1.56	60S ribosomal protein L4 [Canis lupus familiaris]	
gi 187609314	1.54	Chain p, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 74010186	1.52	PREDICTED: pancreatic alpha-amylase [Canis lupus familiaris]	
gi 356582485	1.49	40S ribosomal protein S27 [Canis lupus familiaris]	
gi 545502625	1.46	PREDICTED: 40S ribosomal protein S2 isoform X1 [Canis lupus familiaris]	

gi 57086947	1.46	PREDICTED: 60S ribosomal protein L13 [Canis lupus familiaris]	
gi 345799858	1.43	PREDICTED: 40S ribosomal protein S25 [Canis lupus familiaris]	
gi 73956649	1.29	PREDICTED: 60S ribosomal protein L22 [Canis lupus familiaris]	
gi 89573885	1.28	ribosomal protein L18, partial [Canis lupus familiaris]	
gi 187609261	1.27	Chain h, Structure Of A Mammalian Ribosomal 40s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 545535830	1.25	PREDICTED: 60S ribosomal protein L27a isoform X1 [Canis lupus familiaris]	
gi 5441527	1.23	ribosomal protein L27 [Canis lupus familiaris]	
gi 545503787	1.19	PREDICTED: 60S ribosomal protein L5 [Canis lupus familiaris]	
gi 356460872	1.15	40S ribosomal protein S6 [Canis lupus familiaris]	
gi 73957688	1.13	PREDICTED: transducin beta-like protein 2 [Canis lupus familiaris]	
gi 928184673	1.12	PREDICTED: 60S ribosomal protein L10 isoform X1 [Canis lupus familiaris]	
gi 356460978	1.12	60S ribosomal protein L10a [Canis lupus familiaris]	
gi 928143153	1.10	PREDICTED: 60S ribosomal protein L38 [Canis lupus familiaris]	
gi 57109594	1.06	PREDICTED: 60S ribosomal protein L35a [Canis lupus familiaris]	
gi 136406	1.02	RecName: Full=Cationic trypsin; Flags: Precursor	
gi 545557165	1.02	PREDICTED: translocon-associated protein subunit delta isoform X1 [Canis lupus familiaris]	TRAPδ
gi 545543878	0.97	PREDICTED: 60S ribosomal protein L6 isoform X1 [Canis lupus familiaris]	
gi 187609256	0.96	Chain b, Structure Of A Mammalian Ribosomal 40s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 545542114	0.96	PREDICTED: 60S ribosomal protein L34 [Canis lupus familiaris]	
gi 34979905	0.94	epithelial keratin 10 [Canis lupus familiaris]	
gi 354725914	0.88	40S ribosomal protein S4, X isoform [Canis lupus familiaris]	
gi 187609321	0.88	Chain v, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 73963665	0.87	PREDICTED: transmembrane emp24 domain-containing protein 10 [Canis lupus familiaris]	
gi 356640165	0.84	40S ribosomal protein S24 isoform 1 [Canis lupus familiaris]	
gi 459746	0.82	protein translocation complex beta subunit [Canis lupus familiaris]	Sec61β
gi 73990466	0.81	PREDICTED: translocon-associated protein subunit gamma [Canis lupus familiaris]	TRAPγ
gi 5441541	0.81	Ribosomal protein [Canis lupus familiaris]	
gi 345799231	0.79	PREDICTED: annexin A11 [Canis lupus familiaris]	
gi 545487328	0.79	PREDICTED: 40S ribosomal protein S9 [Canis lupus familiaris]	
gi 187609265	0.78	Chain l, Structure Of A Mammalian Ribosomal 40s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 57048082	0.76	PREDICTED: 60S ribosomal protein L9 [Canis lupus familiaris]	
gi 187609327	0.76	Chain z, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 172046690	0.73	RecName: Full=DnaJ homolog subfamily B member 11; AltName: Full=ER-associated DNAJ; AltName: Full=ER-associated Hsp40 co-chaperone;	
gi 187609300	0.73	Chain b, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 560891938	0.71	ATP synthase subunit alpha, mitochondrial [Canis lupus familiaris]	
gi 73991617	0.71	PREDICTED: 60S ribosomal protein L26 [Canis lupus familiaris]	
gi 545539003	0.69	PREDICTED: signal recognition particle receptor subunit beta [Canis lupus familiaris]	SRβ
gi 187609318	0.69	Chain s, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gi 73966629	0.68	PREDICTED: clathrin heavy chain 1 isoform X2 [Canis lupus familiaris]	
gi 545498051	0.61	PREDICTED: vesicle-associated membrane protein 2 [Canis lupus familiaris]	
gi 73962567	0.61	PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1 [Canis lupus familiaris]	DAD1
gi 73982080	0.58	PREDICTED: very-long-chain 3-oxoacyl-CoA reductase [Canis lupus familiaris]	

gij187609302	0.58	Chain d, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gij34979907	0.57	epithelial keratin 1 [Canis lupus familiaris]	
gij359322142	0.54	PREDICTED: elongation factor 2 [Canis lupus familiaris]	
gij545500190	0.54	PREDICTED: chymotrypsin-like protease CTRL-1 isoform X2 [Canis lupus familiaris]	
gij187609313	0.54	Chain o, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gij74003143	0.54	PREDICTED: programmed cell death protein 6 isoform X1 [Canis lupus familiaris]	
gij359322420	0.53	PREDICTED: eukaryotic translation initiation factor 3 subunit F [Canis lupus familiaris]	
gij74003749	0.53	PREDICTED: 60S ribosomal protein L22-like 1 [Canis lupus familiaris]	
gij187609324	0.53	Chain x, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gij545530860	0.53	PREDICTED: ubiquitin-like protein fubi and ribosomal protein S30 [Canis lupus familiaris]	
gij73991908	0.52	PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2 isoform X2 [Canis lupus familiaris]	RPN2
gij73963693	0.52	PREDICTED: 40S ribosomal protein S29 [Canis lupus familiaris]	
gij57036143	0.50	PREDICTED: 40S ribosomal protein S5 [Canis lupus familiaris]	
gij356991221	0.50	60S ribosomal protein L14 [Canis lupus familiaris]	
gij325530312	0.49	RecName: Full=60S ribosomal protein L13a	
gij44888810	0.49	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	
gij359320346	0.48	PREDICTED: CDK5 regulatory subunit-associated protein 3 [Canis lupus familiaris]	
gij345790954	0.48	PREDICTED: malectin isoform X1 [Canis lupus familiaris]	Malectin
gij73971644	0.48	PREDICTED: ras-related protein Rab-14 [Canis lupus familiaris]	
gij131797	0.48	RecName: Full=Ras-related protein Rab-7a	
gij73989982	0.47	PREDICTED: 60S ribosomal protein L15 isoform X1 [Canis lupus familiaris]	
gij187609325	0.47	Chain y, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gij924442847	0.46	actin, cytoplasmic 2 [Canis lupus familiaris]	
gij848	0.46	glycoprotein 25L [Canis lupus familiaris]	
gij73958588	0.44	PREDICTED: eukaryotic translation initiation factor 3 subunit C [Canis lupus familiaris]	
gij129505	0.44	RecName: Full=Phospholipase A2; AltName: Full=Group IB phospholipase A2; AltName: Full=Phosphatidylcholine 2-acylhydrolase 1B; Flags: Precursor	
gij545541162	0.44	PREDICTED: RNA-binding protein 38 [Canis lupus familiaris]	
gij545496900	0.43	PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A [Canis lupus familiaris]	STT3A
gij73969188	0.42	PREDICTED: eukaryotic translation initiation factor 3 subunit D [Canis lupus familiaris]	
gij136411	0.42	RecName: Full=Anionic trypsin; Flags: Precursor	
gij73968432	0.41	PREDICTED: ATP synthase subunit beta, mitochondrial [Canis lupus familiaris]	
gij545537816	0.41	PREDICTED: dnaJ homolog subfamily C member 3 [Canis lupus familiaris]	
gij354721202	0.39	40S ribosomal protein S10 [Canis lupus familiaris]	
gij164038	0.38	oligosaccharyltransferase 48 kDa subunit [Canis lupus familiaris]	OST48
gij73950876	0.38	PREDICTED: chymotrypsin-C [Canis lupus familiaris]	
gij345783280	0.38	PREDICTED: transmembrane protein 258 [Canis lupus familiaris]	TMEM258
gij73983414	0.37	PREDICTED: elongation factor 1-gamma [Canis lupus familiaris]	
gij73974130	0.37	PREDICTED: polyadenylate-binding protein 1 [Canis lupus familiaris]	
gij73998882	0.36	PREDICTED: pancreatic triacylglycerol lipase [Canis lupus familiaris]	
gij57085107	0.35	PREDICTED: annexin A7 isoform X1 [Canis lupus familiaris]	
gij935	0.35	unnamed protein product [Canis lupus familiaris]	
gij545526854	0.33	PREDICTED: 40S ribosomal protein S7 [Canis lupus familiaris]	
gij57095966	0.33	PREDICTED: sorcin isoform X2 [Canis lupus familiaris]	

gij187609305	0.32	Chain g, Structure Of A Mammalian Ribosomal 60s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The Rna And Proteins Into An 8.7 A Cryo-em Map	
gij915	0.32	rab10 [Canis lupus familiaris]	
gij341865609	0.31	pancreatic secretory granule membrane major glycoprotein GP2 precursor [Canis lupus familiaris]	
gij545555828	0.31	PREDICTED: NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 3 [Canis lupus familiaris]	
gij5731788	0.31	phospholipid hydroperoxide glutathione peroxidase, partial [Canis lupus familiaris]	
gij73980394	0.30	PREDICTED: protein disulfide-isomerase A6 [Canis lupus familiaris]	
gij74000476	0.30	PREDICTED: peptidyl-prolyl cis-trans isomerase B [Canis lupus familiaris]	
gij345780758	0.30	PREDICTED: nuclease-sensitive element-binding protein 1 isoform X1 [Canis lupus familiaris]	
gij73983454	0.30	PREDICTED: 60S acidic ribosomal protein P2 [Canis lupus familiaris]	
gij359320835	0.30	PREDICTED: clathrin light chain A isoform X4 [Canis lupus familiaris]	
gij73973426	0.30	PREDICTED: BAG family molecular chaperone regulator 2 [Canis lupus familiaris]	
gij359320857	0.29	PREDICTED: E3 UFM1-protein ligase 1 [Canis lupus familiaris]	
gij928152172	0.29	PREDICTED: LOW QUALITY PROTEIN: 60S acidic ribosomal protein P1-like [Canis lupus familiaris]	
gij74008235	0.29	PREDICTED: uncharacterized protein LOC480987 [Canis lupus familiaris]	
gij73971210	0.28	PREDICTED: transitional endoplasmic reticulum ATPase [Canis lupus familiaris]	
gij16607669	0.28	unnamed protein product [Canis lupus familiaris]	
gij359323342	0.28	PREDICTED: protein LYRIC isoform X3 [Canis lupus familiaris]	
gij73957259	0.27	PREDICTED: ran-binding protein 10 [Canis lupus familiaris]	
gij73998665	0.27	PREDICTED: eukaryotic translation initiation factor 3 subunit A isoform X1 [Canis lupus familiaris]	
gij73974299	0.27	PREDICTED: eukaryotic translation initiation factor 3 subunit E [Canis lupus familiaris]	
gij17298186	0.26	heat shock protein 70 [Canis lupus familiaris]	
gij928168898	0.26	PREDICTED: eukaryotic translation initiation factor 6 isoform X2 [Canis lupus familiaris]	
gij57092393	0.26	PREDICTED: 40S ribosomal protein S26 [Canis lupus familiaris]	
gij545523499	0.25	PREDICTED: polyadenylate-binding protein 4 isoform X1 [Canis lupus familiaris]	
gij73969600	0.25	PREDICTED: endoplasmic reticulum lectin 1 isoform X2 [Canis lupus familiaris]	
gij73957446	0.25	PREDICTED: transmembrane emp24 domain-containing protein 6 isoform X1 [Canis lupus familiaris]	
gij345800374	0.25	PREDICTED: plasminogen activator inhibitor 1 RNA-binding protein isoform X4 [Canis lupus familiaris]	
gij928136876	0.24	PREDICTED: eukaryotic translation initiation factor 3 subunit B [Canis lupus familiaris]	
gij359279916	0.24	keratin, type I cytoskeletal 19 [Canis lupus familiaris]	
gij359324135	0.24	PREDICTED: peroxiredoxin-4 isoform X4 [Canis lupus familiaris]	
gij158442066	0.24	prohibitin [Canis lupus familiaris]	
gij928143386	0.23	PREDICTED: E2/E3 hybrid ubiquitin-protein ligase UBE2O [Canis lupus familiaris]	
gij73990231	0.23	PREDICTED: armadillo repeat-containing protein 8 isoform X2 [Canis lupus familiaris]	
gij73980965	0.23	PREDICTED: mannosyl-oligosaccharide glucosidase [Canis lupus familiaris]	
gij545522024	0.23	PREDICTED: carboxypeptidase A1 [Canis lupus familiaris]	
gij226732647	0.23	RecName: Full=Keratinocyte-associated protein 2; Short=KCP-2	KCP2
gij73948370	0.23	PREDICTED: syncollin [Canis lupus familiaris]	
gij354459407	0.23	voltage-dependent anion-selective channel protein 3 isoform 2 [Canis lupus familiaris]	
gij74009006	0.23	PREDICTED: membrane magnesium transporter 1 [Canis lupus familiaris]	EMC5
gij928147174	0.23	PREDICTED: NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6 [Canis lupus familiaris]	
gij57087967	0.22	PREDICTED: cytochrome b-c1 complex subunit 2, mitochondrial [Canis lupus familiaris]	
gij73975552	0.22	PREDICTED: carboxypeptidase A2 [Canis lupus familiaris]	
gij16975342	0.22	Chain A, Crystal Structure Of The Lumenal Domain Of Calnexin	
gij34979903	0.22	epithelial keratin 2e [Canis lupus familiaris]	

gi 545492518	0.22	PREDICTED: receptor expression-enhancing protein 5, partial [Canis lupus familiaris]	
gi 40313536	0.22	deoxyribonuclease I [Canis lupus familiaris]	
gi 73983273	0.22	PREDICTED: peptidyl-prolyl cis-trans isomerase FKBP2 isoform X2 [Canis lupus familiaris]	
gi 345785949	0.22	PREDICTED: 60S ribosomal protein L28 [Canis lupus familiaris]	
gi 359323129	0.21	PREDICTED: tubulin alpha-1C chain [Canis lupus familiaris]	
gi 308199425	0.21	elongation factor 1-alpha 1 [Canis lupus familiaris]	
gi 73979580	0.21	PREDICTED: ADP/ATP translocase 1 [Canis lupus familiaris]	
gi 74008194	0.21	PREDICTED: ADP/ATP translocase 2 [Canis lupus familiaris]	
gi 545541681	0.21	PREDICTED: ubiquitin-fold modifier 1 isoform X1 [Canis lupus familiaris]	
gi 545521108	0.20	PREDICTED: elongation factor 1-delta isoform X1 [Canis lupus familiaris]	
gi 206984847	0.20	unnamed protein product [Canis lupus familiaris]	
gi 126316	0.20	RecName: Full=Inactive pancreatic lipase-related protein 1; Short=PL-RP1; Flags: Precursor	
gi 226739213	0.20	RecName: Full=Oligosaccharyltransferase complex subunit OSTC	DC2/OSTC
gi 545534022	0.20	PREDICTED: NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13 [Canis lupus familiaris]	
gi 73956140	0.19	PREDICTED: fatty aldehyde dehydrogenase isoform X2 [Canis lupus familiaris]	
gi 545503669	0.19	PREDICTED: nodal modulator 1 [Canis lupus familiaris]	
gi 928128521	0.19	PREDICTED: eukaryotic translation initiation factor 3 subunit I [Canis lupus familiaris]	
gi 928183240	0.19	PREDICTED: transmembrane and coiled-coil domain-containing protein 1 isoform X1 [Canis lupus familiaris]	TMCO1
gi 359319817	0.18	PREDICTED: protein disulfide-isomerase A2 isoform X2 [Canis lupus familiaris]	
gi 73979213	0.18	PREDICTED: erlin-2 isoform X2 [Canis lupus familiaris]	
gi 545501397	0.18	PREDICTED: vitamin K epoxide reductase complex subunit 1 [Canis lupus familiaris]	
gi 57093423	0.18	PREDICTED: poly(rC)-binding protein 1 [Canis lupus familiaris]	
gi 74003099	0.18	PREDICTED: ubiquitin-conjugating enzyme E2Q-like protein 1 [Canis lupus familiaris]	
gi 928167632	0.17	PREDICTED: serotransferrin [Canis lupus familiaris]	
gi 1333622	0.17	paraoxonase 2 [Canis lupus familiaris]	
gi 57111763	0.17	PREDICTED: PRA1 family protein 2 [Canis lupus familiaris]	
gi 475062	0.17	VIP36 (vesicular integral-membrane protein) [Canis lupus familiaris]	
gi 345791839	0.16	PREDICTED: keratin, type II cytoskeletal 6A isoform X2 [Canis lupus familiaris]	
gi 10946310	0.16	transferrin receptor [Canis lupus familiaris]	
gi 545514391	0.16	PREDICTED: cysteine-rich with EGF-like domain protein 2 [Canis lupus familiaris]	
gi 345807587	0.16	PREDICTED: magnesium transporter protein 1 isoform X1 [Canis lupus familiaris]	MAGT1
gi 134792	0.16	RecName: Full=Signal peptidase complex catalytic subunit SEC11C; AltName: Full=Microsomal signal peptidase 21 kDa subunit; Short=SPase 21 kDa subunit;	Sec11
gi 73998777	0.15	PREDICTED: CUB and zona pellucida-like domain-containing protein 1 [Canis lupus familiaris]	
gi 345799449	0.15	PREDICTED: clathrin interactor 1 isoform X1 [Canis lupus familiaris]	
gi 73989883	0.15	PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B [Canis lupus familiaris]	STT3B
gi 353731063	0.15	eukaryotic translation initiation factor 3 subunit L [Canis lupus familiaris]	
gi 345796543	0.15	PREDICTED: translocation protein SEC62 [Canis lupus familiaris]	Sec62
gi 345777399	0.15	PREDICTED: AP-3 complex subunit sigma-1 isoform X1 [Canis lupus familiaris]	
gi 73980315	0.15	PREDICTED: protein RMD5 homolog A [Canis lupus familiaris]	
gi 73985044	0.15	PREDICTED: ADP-ribosylation factor-like protein 8B [Canis lupus familiaris]	
gi 345791445	0.14	PREDICTED: stromal cell-derived factor 2-like protein 1 [Canis lupus familiaris]	
gi 356460984	0.14	eukaryotic peptide chain release factor subunit 1 [Canis lupus familiaris]	
gi 415377	0.14	Rab5c protein [Canis lupus familiaris]	
gi 545551111	0.14	PREDICTED: ATP synthase subunit O, mitochondrial [Canis lupus familiaris]	
gi 57106334	0.14	PREDICTED: keratin, type I cytoskeletal 18 [Canis lupus familiaris]	
gi 57110955	0.13	PREDICTED: elongation factor 1-beta [Canis lupus familiaris]	
gi 545533839	0.13	PREDICTED: cytochrome b-c1 complex subunit 1, mitochondrial isoform X1 [Canis lupus familiaris]	

gi 57110953	0.13	PREDICTED: NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial [Canis lupus familiaris]	
gi 57098751	0.13	PREDICTED: vesicle-trafficking protein SEC22b [Canis lupus familiaris]	
gi 359320233	0.13	PREDICTED: serine/arginine-rich splicing factor 2 [Canis lupus familiaris]	
gi 164070	0.13	homologue to sec61 [Canis lupus familiaris]	Sec61 α
gi 57036583	0.13	PREDICTED: eukaryotic translation initiation factor 3 subunit K [Canis lupus familiaris]	
gi 533111	0.13	signal peptidase complex 25 kDa subunit [Canis lupus familiaris]	SPase 25 kDa
gi 545549261	0.13	PREDICTED: ER membrane protein complex subunit 7 [Canis lupus familiaris]	EMC7
gi 73993818	0.12	PREDICTED: V-type proton ATPase subunit B, brain isoform [Canis lupus familiaris]	
gi 345779293	0.12	PREDICTED: transmembrane protein 33 [Canis lupus familiaris]	
gi 73945930	0.12	PREDICTED: protein ERGIC-53 isoform X1 [Canis lupus familiaris]	
gi 374532808	0.12	thioredoxin-dependent peroxide reductase, mitochondrial [Canis lupus familiaris]	
gi 345794865	0.12	PREDICTED: protein disulfide-isomerase A3 [Canis lupus familiaris]	
gi 117612	0.12	RecName: Full=Chymotrypsinogen 2; Contains: RecName: Full=Chymotrypsin 2 chain A; Contains: RecName: Full=Chymotrypsin 2 chain B; Flags: Precursor	
gi 73980752	0.12	PREDICTED: serine/arginine-rich splicing factor 7 isoform X1 [Canis lupus familiaris]	
gi 73992430	0.12	PREDICTED: glucose-induced degradation protein 8 homolog [Canis lupus familiaris]	
gi 73946982	0.12	PREDICTED: ER membrane protein complex subunit 10 isoform X2 [Canis lupus familiaris]	EMC10
gi 73962198	0.12	PREDICTED: NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial [Canis lupus familiaris]	
gi 73964749	0.12	PREDICTED: protein disulfide-isomerase [Canis lupus familiaris]	
gi 73992635	0.12	PREDICTED: vesicle-associated membrane protein-associated protein B/C isoform X2 [Canis lupus familiaris]	
gi 345792055	0.12	PREDICTED: methyltransferase-like protein 7A [Canis lupus familiaris]	
gi 345777664	0.12	PREDICTED: protein NipSnap homolog 3A [Canis lupus familiaris]	
gi 345800975	0.11	PREDICTED: enhancer of mRNA-decapping protein 4 [Canis lupus familiaris]	
gi 345791833	0.11	PREDICTED: keratin, type II cytoskeletal 3 [Canis lupus familiaris]	
gi 57043329	0.11	PREDICTED: proproteinase E [Canis lupus familiaris]	
gi 928180988	0.11	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC488260 [Canis lupus familiaris]	
gi 73983298	0.11	PREDICTED: NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial [Canis lupus familiaris]	
gi 57100904	0.11	PREDICTED: ER membrane protein complex subunit 3 [Canis lupus familiaris]	EMC3
gi 545525725	0.11	PREDICTED: transmembrane and ubiquitin-like domain-containing protein 1 [Canis lupus familiaris]	
gi 73986020	0.10	PREDICTED: manganese-transporting ATPase 13A1 [Canis lupus familiaris]	
gi 134891	0.10	RecName: Full=Signal recognition particle receptor subunit alpha; Short=SR-alpha; AltName: Full=Docking protein alpha; Short=DP-alpha	SR α
gi 345780013	0.10	PREDICTED: staphylococcal nuclease domain-containing protein 1 [Canis lupus familiaris]	
gi 73953093	0.10	PREDICTED: voltage-dependent anion-selective channel protein 2 [Canis lupus familiaris]	
gi 74006309	0.10	PREDICTED: DDB1- and CUL4-associated factor 8 [Canis lupus familiaris]	
gi 57109638	0.10	PREDICTED: V-type proton ATPase catalytic subunit A [Canis lupus familiaris]	
gi 928161789	0.10	PREDICTED: membrane-associated progesterone receptor component 2, partial [Canis lupus familiaris]	
gi 928159936	0.10	PREDICTED: AP-2 complex subunit alpha-2 [Canis lupus familiaris]	
gi 73965738	0.10	PREDICTED: V-type proton ATPase 116 kDa subunit a isoform 1 isoform X6 [Canis lupus familiaris]	
gi 345791621	0.10	PREDICTED: prohibitin-2 [Canis lupus familiaris]	
gi 2853285	0.09	sarcoplasmic reticulum Ca ²⁺ -transport ATPase isoform [Canis lupus familiaris]	
gi 359321459	0.09	PREDICTED: protein disulfide-isomerase A4 [Canis lupus familiaris]	
gi 73960960	0.09	PREDICTED: secretory carrier-associated membrane protein 3 isoform X1 [Canis lupus familiaris]	
gi 545513876	0.09	PREDICTED: protein OS-9 isoform X1 [Canis lupus familiaris]	

gi 73974479	0.09	PREDICTED: ER membrane protein complex subunit 2 [Canis lupus familiaris]	EMC2
gi 73978790	0.09	PREDICTED: dnaJ homolog subfamily B member 6 isoform X1 [Canis lupus familiaris]	
gi 73986278	0.09	PREDICTED: dnaJ homolog subfamily B member 1 isoform X1 [Canis lupus familiaris]	
gi 545537385	0.08	PREDICTED: fibronectin type-III domain-containing protein 3A isoform X1 [Canis lupus familiaris]	
gi 73983859	0.08	PREDICTED: DNA damage-binding protein 1 [Canis lupus familiaris]	
gi 928176123	0.08	PREDICTED: aspartyl/asparaginyl beta-hydroxylase isoform X2 [Canis lupus familiaris]	
gi 928157470	0.08	PREDICTED: transmembrane protein 214 [Canis lupus familiaris]	
gi 73957263	0.08	PREDICTED: V-type proton ATPase subunit d 1 [Canis lupus familiaris]	
gi 345794639	0.08	PREDICTED: thrombospondin-1 [Canis lupus familiaris]	
gi 73979498	0.08	PREDICTED: long-chain-fatty-acid--CoA ligase 1 isoform X1 [Canis lupus familiaris]	
gi 24021296	0.08	glutamine synthetase [Canis lupus familiaris]	
gi 359322750	0.08	PREDICTED: serine/arginine-rich splicing factor 6 isoform X1 [Canis lupus familiaris]	
gi 73991533	0.08	PREDICTED: minor histocompatibility antigen H13 isoform X4 [Canis lupus familiaris]	
gi 942	0.08	TRAM-protein [Canis lupus familiaris]	TRAM
gi 219772678	0.07	unnamed protein product [Canis lupus familiaris]	
gi 359320880	0.07	PREDICTED: valine--tRNA ligase isoform X2 [Canis lupus familiaris]	
gi 928130713	0.07	PREDICTED: macrophage erythroblast attacher isoform X1 [Canis lupus familiaris]	
gi 345796421	0.07	PREDICTED: eukaryotic initiation factor 4A-II [Canis lupus familiaris]	
gi 545488118	0.07	PREDICTED: glutaminyl-peptide cyclotransferase-like protein isoform X2 [Canis lupus familiaris]	
gi 73979922	0.07	PREDICTED: prolactin regulatory element-binding protein [Canis lupus familiaris]	
gi 545531873	0.06	PREDICTED: UDP-glucose:glycoprotein glucosyltransferase 1 [Canis lupus familiaris]	
gi 345793750	0.06	PREDICTED: ER membrane protein complex subunit 1 isoform X6 [Canis lupus familiaris]	EMC1
gi 359318595	0.06	PREDICTED: AP-2 complex subunit alpha-1 isoform X1 [Canis lupus familiaris]	
gi 73960479	0.06	PREDICTED: ER degradation-enhancing alpha-mannosidase-like protein 3 isoform X7 [Canis lupus familiaris]	
gi 928131403	0.06	PREDICTED: glutamate dehydrogenase 1, mitochondrial, partial [Canis lupus familiaris]	
gi 545499442	0.06	PREDICTED: tubulin beta-3 chain [Canis lupus familiaris]	
gi 545521446	0.06	PREDICTED: NF-X1-type zinc finger protein NFXL1 [Canis lupus familiaris]	
gi 930	0.06	unnamed protein product [Canis lupus familiaris]	
gi 74009203	0.06	PREDICTED: H/ACA ribonucleoprotein complex subunit 4 [Canis lupus familiaris]	
gi 359323340	0.06	PREDICTED: phosphatidylserine synthase 1 isoform X1 [Canis lupus familiaris]	
gi 345784138	0.06	PREDICTED: metalloredutase STEAP3 isoform X2 [Canis lupus familiaris]	
gi 73979871	0.06	PREDICTED: trifunctional enzyme subunit beta, mitochondrial [Canis lupus familiaris]	
gi 928142969	0.05	PREDICTED: heat shock protein HSP 90-alpha [Canis lupus familiaris]	
gi 928181018	0.05	PREDICTED: LOW QUALITY PROTEIN: tripartite motif-containing protein 10 [Canis lupus familiaris]	
gi 57098645	0.05	PREDICTED: all-trans-retinol 13,14-reductase isoform X1 [Canis lupus familiaris]	
gi 345793181	0.05	PREDICTED: copine-3 [Canis lupus familiaris]	
gi 345781155	0.05	PREDICTED: ATP-binding cassette sub-family E member 1 isoform X1 [Canis lupus familiaris]	
gi 57092243	0.04	PREDICTED: bile salt-activated lipase [Canis lupus familiaris]	
gi 403497	0.04	GRP94 [Canis lupus familiaris]	
gi 74003954	0.04	PREDICTED: phosphatidylinositol-glycan-specific phospholipase D isoform X2 [Canis lupus familiaris]	
gi 73980337	0.04	PREDICTED: MICOS complex subunit MIC60 isoform X1 [Canis lupus familiaris]	
gi 359320614	0.04	PREDICTED: LOW QUALITY PROTEIN: lipase maturation factor 2 [Canis lupus familiaris]	
gi 73964368	0.04	PREDICTED: protein sel-1 homolog 1 isoform X1 [Canis lupus familiaris]	
gi 545500961	0.04	PREDICTED: pentatricopeptide repeat-containing protein 1, mitochondrial isoform X4 [Canis lupus familiaris]	
gi 359318618	0.04	PREDICTED: LOW QUALITY PROTEIN: glycogen [starch] synthase, muscle [Canis lupus familiaris]	

gi 545556290	0.03	PREDICTED: ATP-binding cassette sub-family B member 6, mitochondrial [Canis lupus familiaris]	
gi 33413904	0.03	adaptor-related protein complex AP3 beta 1 subunit [Canis lupus familiaris]	
gi 73982653	0.03	PREDICTED: cysteine--tRNA ligase, cytoplasmic isoform X1 [Canis lupus familiaris]	
gi 73966186	0.03	PREDICTED: importin subunit beta-1 [Canis lupus familiaris]	
gi 545495990	0.03	PREDICTED: Ia-related protein 1 [Canis lupus familiaris]	
gi 73964198	0.03	PREDICTED: C-1-tetrahydrofolate synthase, cytoplasmic isoform X1 [Canis lupus familiaris]	
gi 545503813	0.03	PREDICTED: glutamate-rich protein 3 [Canis lupus familiaris]	
gi 928146436	0.02	PREDICTED: extended synaptotagmin-1 [Canis lupus familiaris]	
gi 545543200	0.02	PREDICTED: vigilin [Canis lupus familiaris]	
gi 928156856	0.02	PREDICTED: ankyrin-1 isoform X1 [Canis lupus familiaris]	
gi 359319316	0.02	PREDICTED: zinc finger SWIM domain-containing protein 8 isoform X3 [Canis lupus familiaris]	
gi 928129726	0.02	PREDICTED: putative ankyrin repeat domain-containing protein 31 isoform X1 [Canis lupus familiaris]	
gi 333805632	0.01	spectrin beta chain, erythrocytic [Canis lupus familiaris]	

A.2 Hits from the *oryctolagus cuniculus* database

Table A.2 – MS analysis of the solubilized OP96 sample - Hits from *oryctolagus cuniculus* (rabbit). Mass spectrometry analysis was performed by Thomas Fröhlich (Gene Center, LMU), hits are sorted according to their abundance as quantified by the emPAI value. The database from *oryctolagus cuniculus* was used for the identification of soluble components due to the use of rabbit reticulocyte lysate in the translation reaction. Hits from membrane components arise from high similarity between *canis lupus familiaris* (PKRM) and *oryctolagus cuniculus* (RRL).

Accession	emPAI	Description
gi 529281378	8.19	Chain c, Rabbit 40s Ribosomal Subunit In Complex With Eif1.
gi 291390048	7.47	PREDICTED: 60S ribosomal protein L38 [Oryctolagus cuniculus]
gi 665764245	5.76	Chain T, Regulation Of The Mammalian Elongation Cycle By 40s Subunit Rolling: A Eukaryotic-specific Ribosome Rearrangement
gi 291388327	5.49	PREDICTED: 60S ribosomal protein L30 [Oryctolagus cuniculus]
gi 291396021	4.70	PREDICTED: 40S ribosomal protein S18 [Oryctolagus cuniculus]
gi 291400758	4.64	PREDICTED: 60S ribosomal protein L24 [Oryctolagus cuniculus]
gi 665764286	4.14	Chain c, Regulation Of The Mammalian Elongation Cycle By 40s Subunit Rolling: A Eukaryotic-specific Ribosome Rearrangement
gi 665764247	3.98	Chain V, Regulation Of The Mammalian Elongation Cycle By 40s Subunit Rolling: A Eukaryotic-specific Ribosome Rearrangement
gi 291389960	3.80	PREDICTED: 40S ribosomal protein S3a [Oryctolagus cuniculus]
gi 291416370	3.58	PREDICTED: ubiquitin-40S ribosomal protein S27a [Oryctolagus cuniculus]
gi 1485	3.57	unnamed protein product [Oryctolagus cuniculus]
gi 291408313	2.96	PREDICTED: 60S ribosomal protein L12 [Oryctolagus cuniculus]
gi 89573875	2.75	ribosomal protein L18, partial [Oryctolagus cuniculus]
gi 291387632	2.71	PREDICTED: 40S ribosomal protein S14 [Oryctolagus cuniculus]
gi 291394359	2.66	PREDICTED: 60S ribosomal protein L17 [Oryctolagus cuniculus]
gi 291397006	2.44	PREDICTED: 40S ribosomal protein S12 [Oryctolagus cuniculus]
gi 291414539	2.42	PREDICTED: 60S ribosomal protein L8 [Oryctolagus cuniculus]
gi 291400487	2.31	PREDICTED: transferrin receptor protein 1 [Oryctolagus cuniculus]
gi 291384249	2.22	PREDICTED: 40S ribosomal protein S3 [Oryctolagus cuniculus]
gi 291386783	2.22	PREDICTED: 40S ribosomal protein S11 [Oryctolagus cuniculus]
gi 291388155	2.20	PREDICTED: 60S ribosomal protein L21 [Oryctolagus cuniculus]
gi 529281382	2.19	Chain g, Rabbit 40s Ribosomal Subunit In Complex With Eif1.
gi 291383687	2.17	PREDICTED: 40S ribosomal protein S8 [Oryctolagus cuniculus]
gi 665764265	2.16	Chain G, Regulation Of The Mammalian Elongation Cycle By 40s Subunit Rolling: A Eukaryotic-specific Ribosome Rearrangement
gi 1483	2.12	beta-globin [Oryctolagus cuniculus]
gi 291389987	2.06	PREDICTED: 40S ribosomal protein S16 [Oryctolagus cuniculus]
gi 655878299	2.03	PREDICTED: 78 kDa glucose-regulated protein [Oryctolagus cuniculus]
gi 307775447	1.91	60S ribosomal protein L4 [Oryctolagus cuniculus]
gi 1751	1.89	liver transferrin [Oryctolagus cuniculus]
gi 665764276	1.83	Chain S, Regulation Of The Mammalian Elongation Cycle By 40s Subunit Rolling: A Eukaryotic-specific Ribosome Rearrangement
gi 291388052	1.80	PREDICTED: 60S ribosomal protein L13 [Oryctolagus cuniculus]
gi 655870993	1.75	PREDICTED: 60S ribosomal protein L23 [Oryctolagus cuniculus]
gi 291403978	1.73	PREDICTED: 60S ribosomal protein L5 [Oryctolagus cuniculus]
gi 291383777	1.70	PREDICTED: heat shock cognate 71 kDa protein [Oryctolagus cuniculus]
gi 365177544	1.67	alpha-globin 1 [Oryctolagus cuniculus]
gi 122475	1.67	RecName: Full=Hemoglobin subunit alpha-1/2; AltName: Full=Alpha-1/2-globin; AltName: Full=Hemoglobin alpha-1/2 chain
gi 291388125	1.64	PREDICTED: 60S ribosomal protein L7 [Oryctolagus cuniculus]
gi 665764275	1.54	Chain R, Regulation Of The Mammalian Elongation Cycle By 40s Subunit Rolling: A Eukaryotic-specific Ribosome Rearrangement
gi 291383303	1.48	PREDICTED: 40S ribosomal protein S20 [Oryctolagus cuniculus]
gi 291410235	1.45	PREDICTED: 40S ribosomal protein S13 [Oryctolagus cuniculus]
gi 291412982	1.43	PREDICTED: 40S ribosomal protein S25 [Oryctolagus cuniculus]

gi 461517	1.42	RecName: Full=Annexin A11; AltName: Full=Annexin XI; AltName: Full=Annexin-11;
gi 217418273	1.41	ribosomal protein L10 (predicted) [Oryctolagus cuniculus]
gi 291383209	1.40	PREDICTED: 40S ribosomal protein S6 [Oryctolagus cuniculus]
gi 291407689	1.33	PREDICTED: 40S ribosomal protein S4, X isoform [Oryctolagus cuniculus]
gi 665764278	1.28	Chain U, Regulation Of The Mammalian Elongation Cycle By 40s Subunit Rolling: A Eukaryotic-specific Ribosome Rearrangement
gi 529281372	1.27	Chain W, Rabbit 40s Ribosomal Subunit In Complex With Eif1.
gi 217418260	1.24	signal sequence receptor, delta (predicted) [Oryctolagus cuniculus]
gi 665764228	1.22	Chain C, Regulation Of The Mammalian Elongation Cycle By 40s Subunit Rolling: A Eukaryotic-specific Ribosome Rearrangement
gi 291393273	1.20	PREDICTED: 60S ribosomal protein L14 [Oryctolagus cuniculus]
gi 655602744	1.14	PREDICTED: 60S ribosomal protein L27a isoform X1 [Oryctolagus cuniculus]
gi 291396081	1.12	PREDICTED: 60S ribosomal protein L10a [Oryctolagus cuniculus]
gi 665764274	1.05	Chain Q, Regulation Of The Mammalian Elongation Cycle By 40s Subunit Rolling: A Eukaryotic-specific Ribosome Rearrangement
gi 291394855	1.04	PREDICTED: sorcin isoform X1 [Oryctolagus cuniculus]
gi 291385667	1.03	PREDICTED: 60S ribosomal protein L9 [Oryctolagus cuniculus]
gi 291389071	1.00	PREDICTED: peptidyl-prolyl cis-trans isomerase FKBP11 [Oryctolagus cuniculus]
gi 291408688	1.00	PREDICTED: 60S ribosomal protein L18a [Oryctolagus cuniculus]
gi 217418300	0.99	ribosomal protein S9 (predicted) [Oryctolagus cuniculus]
gi 655899014	0.94	PREDICTED: 60S ribosomal protein L13a [Oryctolagus cuniculus]
gi 146674796	0.93	laminin receptor precursor, partial [Oryctolagus cuniculus]
gi 291393348	0.92	PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 [Oryctolagus cuniculus]
gi 655878391	0.88	PREDICTED: 60S ribosomal protein L35 [Oryctolagus cuniculus]
gi 1370279	0.87	transmembrane protein [Oryctolagus cuniculus]
gi 291389862	0.86	PREDICTED: 60S ribosomal protein L3 [Oryctolagus cuniculus]
gi 291383381	0.84	PREDICTED: 40S ribosomal protein S27 [Oryctolagus cuniculus]
gi 291404117	0.83	PREDICTED: 40S ribosomal protein S24 isoform X3 [Oryctolagus cuniculus]
gi 291382907	0.82	PREDICTED: protein transport protein Sec61 subunit beta [Oryctolagus cuniculus]
gi 291400056	0.81	PREDICTED: translocon-associated protein subunit gamma [Oryctolagus cuniculus]
gi 291394851	0.81	PREDICTED: 40S ribosomal protein S17 [Oryctolagus cuniculus]
gi 291395525	0.79	PREDICTED: 60S ribosomal protein L27 [Oryctolagus cuniculus]
gi 291393699	0.79	PREDICTED: 60S ribosomal protein L28 [Oryctolagus cuniculus]
gi 291406014	0.78	PREDICTED: keratin, type I cytoskeletal 10 isoform X1 [Oryctolagus cuniculus]
gi 291406991	0.78	PREDICTED: 60S ribosomal protein L6 [Oryctolagus cuniculus]
gi 291394968	0.78	PREDICTED: 40S ribosomal protein S23 [Oryctolagus cuniculus]
gi 1553	0.77	elongation factor 1 gamma [Oryctolagus cuniculus]
gi 291411482	0.77	PREDICTED: transducin beta-like protein 2 [Oryctolagus cuniculus]
gi 291403291	0.75	PREDICTED: putative 60S ribosomal protein L37a [Oryctolagus cuniculus]
gi 655895819	0.71	PREDICTED: tripartite motif-containing protein 58 [Oryctolagus cuniculus]
gi 291405068	0.71	PREDICTED: 60S ribosomal protein L26 [Oryctolagus cuniculus]
gi 655852411	0.69	PREDICTED: signal recognition particle receptor subunit beta [Oryctolagus cuniculus]
gi 655851870	0.69	PREDICTED: vesicle-associated membrane protein 3 [Oryctolagus cuniculus]
gi 291386554	0.69	PREDICTED: 60S ribosomal protein L23a [Oryctolagus cuniculus]
gi 291395920	0.67	PREDICTED: tripartite motif-containing protein 10 isoform X1 [Oryctolagus cuniculus]
gi 655871198	0.67	PREDICTED: keratin, type I cytoskeletal 15 [Oryctolagus cuniculus]
gi 655870340	0.62	PREDICTED: clathrin heavy chain 1 isoform X1 [Oryctolagus cuniculus]

gi 291383906	0.61	PREDICTED: 60S ribosomal protein L36a [Oryctolagus cuniculus]
gi 327315387	0.61	60S ribosomal protein L35a [Oryctolagus cuniculus]
gi 291400281	0.60	PREDICTED: dnaJ homolog subfamily B member 11 [Oryctolagus cuniculus]
gi 291407056	0.58	PREDICTED: 60S acidic ribosomal protein P0 [Oryctolagus cuniculus]
gi 291403718	0.57	PREDICTED: 60S ribosomal protein L34 [Oryctolagus cuniculus]
gi 291413981	0.53	PREDICTED: programmed cell death protein 6 isoform X1 [Oryctolagus cuniculus]
gi 291386196	0.53	PREDICTED: 60S ribosomal protein L31 [Oryctolagus cuniculus]
gi 89574025	0.52	mitochondrial ATP synthase, H ⁺ transporting F1 complex beta subunit, partial [Oryctolagus cuniculus]
gi 291400180	0.52	PREDICTED: 60S ribosomal protein L22-like 1 [Oryctolagus cuniculus]
gi 291414405	0.52	PREDICTED: ubiquitin-like protein fubi and ribosomal protein S30 [Oryctolagus cuniculus]
gi 291388310	0.52	PREDICTED: 40S ribosomal protein S29 [Oryctolagus cuniculus]
gi 291411397	0.50	PREDICTED: 40S ribosomal protein S5 [Oryctolagus cuniculus]
gi 1203894	0.50	elongation factor 1 delta [Oryctolagus cuniculus]
gi 291408446	0.48	PREDICTED: ras-related protein Rab-14 [Oryctolagus cuniculus]
gi 655870908	0.48	PREDICTED: CDK5 regulatory subunit-associated protein 3 [Oryctolagus cuniculus]
gi 291396428	0.48	PREDICTED: BAG family molecular chaperone regulator 2 [Oryctolagus cuniculus]
gi 291399716	0.47	PREDICTED: 60S ribosomal protein L15 [Oryctolagus cuniculus]
gi 655863889	0.47	PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1 [Oryctolagus cuniculus]
gi 655721756	0.46	PREDICTED: LOW QUALITY PROTEIN: ribosome-binding protein 1 [Oryctolagus cuniculus]
gi 655791666	0.46	PREDICTED: ran-binding protein 10 [Oryctolagus cuniculus]
gi 291390868	0.44	PREDICTED: eukaryotic translation initiation factor 3 subunit C [Oryctolagus cuniculus]
gi 291404166	0.44	PREDICTED: annexin A7 isoform X1 [Oryctolagus cuniculus]
gi 291405799	0.43	PREDICTED: leucine-rich repeat-containing protein 59 [Oryctolagus cuniculus]
gi 291389904	0.42	PREDICTED: eukaryotic translation initiation factor 3 subunit D [Oryctolagus cuniculus]
gi 655893183	0.41	PREDICTED: LOW QUALITY PROTEIN: ubiquitin-conjugating enzyme E2 O [Oryctolagus cuniculus]
gi 655708163	0.41	PREDICTED: eukaryotic translation initiation factor 6 [Oryctolagus cuniculus]
gi 655602713	0.40	PREDICTED: eukaryotic translation initiation factor 3 subunit F [Oryctolagus cuniculus]
gi 291396063	0.39	PREDICTED: 40S ribosomal protein S10 [Oryctolagus cuniculus]
gi 291394323	0.38	PREDICTED: ATP synthase subunit alpha, mitochondrial [Oryctolagus cuniculus]
gi 291383033	0.38	PREDICTED: transitional endoplasmic reticulum ATPase [Oryctolagus cuniculus]
gi 291409694	0.38	PREDICTED: ubiquitin-fold modifier 1 [Oryctolagus cuniculus]
gi 655602634	0.37	PREDICTED: interferon-induced very large GTPase 1-like isoform X1 [Oryctolagus cuniculus]
gi 655688285	0.37	PREDICTED: polyadenylate-binding protein 1 [Oryctolagus cuniculus]
gi 655870053	0.36	PREDICTED: schlafen family member 14 [Oryctolagus cuniculus]
gi 655850868	0.36	PREDICTED: 60S ribosomal protein L11 [Oryctolagus cuniculus]
gi 291407031	0.35	PREDICTED: malectin [Oryctolagus cuniculus]
gi 655852467	0.34	PREDICTED: armadillo repeat-containing protein 8 isoform X1 [Oryctolagus cuniculus]
gi 291397946	0.34	PREDICTED: protein S100-A8 [Oryctolagus cuniculus]
gi 655868457	0.33	PREDICTED: arachidonate 15-lipoxygenase isoform X1 [Oryctolagus cuniculus]
gi 529281357	0.33	Chain H, Rabbit 40s Ribosomal Subunit In Complex With Eif1.
gi 291407802	0.33	PREDICTED: ADP/ATP translocase 2 [Oryctolagus cuniculus]
gi 655869751	0.32	PREDICTED: galectin-9 isoform X1 [Oryctolagus cuniculus]
gi 291383631	0.32	PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A [Oryctolagus cuniculus]
gi 291387085	0.32	PREDICTED: ras-related protein Rab-10 isoform X1 [Oryctolagus cuniculus]

gi 291402876	0.30	PREDICTED: LOW QUALITY PROTEIN: peptidyl-prolyl cis-trans isomerase B [Oryctolagus cuniculus]
gi 608518	0.30	p50 [Oryctolagus cuniculus]
gi 4105817	0.30	Rab7 [Oryctolagus cuniculus]
gi 291382979	0.30	PREDICTED: clathrin light chain A isoform X4 [Oryctolagus cuniculus]
gi 291412313	0.30	PREDICTED: 60S acidic ribosomal protein P1 [Oryctolagus cuniculus]
gi 655606069	0.30	PREDICTED: vesicle-associated membrane protein 8 [Oryctolagus cuniculus]
gi 655881271	0.29	PREDICTED: RNA-binding protein 24 [Oryctolagus cuniculus]
gi 655846321	0.28	PREDICTED: secretory carrier-associated membrane protein 3 [Oryctolagus cuniculus]
gi 291383133	0.28	PREDICTED: 60S ribosomal protein L29-like [Oryctolagus cuniculus]
gi 291384176	0.28	PREDICTED: 60S ribosomal protein L36 [Oryctolagus cuniculus]
gi 655868267	0.27	PREDICTED: eukaryotic translation initiation factor 3 subunit A [Oryctolagus cuniculus]
gi 3108211	0.27	histone H2A.F/Z variant [Oryctolagus cuniculus]
gi 655840227	0.26	PREDICTED: valine--tRNA ligase [Oryctolagus cuniculus]
gi 655842765	0.26	PREDICTED: keratinocyte-associated protein 2-like [Oryctolagus cuniculus]
gi 655736937	0.26	PREDICTED: 40S ribosomal protein S26 [Oryctolagus cuniculus]
gi 291399204	0.25	PREDICTED: polyadenylate-binding protein 4 isoform X2 [Oryctolagus cuniculus]
gi 291398395	0.25	PREDICTED: pancreatic alpha-amylase-like [Oryctolagus cuniculus]
gi 655832614	0.25	PREDICTED: LOW QUALITY PROTEIN: dnaJ homolog subfamily C member 3 [Oryctolagus cuniculus]
gi 291398727	0.25	PREDICTED: plasminogen activator inhibitor 1 RNA-binding protein isoform X4 [Oryctolagus cuniculus]
gi 291396648	0.24	PREDICTED: E3 UFM1-protein ligase 1 isoform X1 [Oryctolagus cuniculus]
gi 655769425	0.24	PREDICTED: cytoskeleton-associated protein 4 [Oryctolagus cuniculus]
gi 217038304	0.24	SEC63-like protein (predicted) [Oryctolagus cuniculus]
gi 291405834	0.24	PREDICTED: prohibitin [Oryctolagus cuniculus]
gi 655899077	0.23	PREDICTED: eukaryotic translation initiation factor 3 subunit B [Oryctolagus cuniculus]
gi 6653665	0.23	voltage-dependent anion channel 3 [Oryctolagus cuniculus]
gi 291404134	0.22	PREDICTED: AP-3 complex subunit mu-1 [Oryctolagus cuniculus]
gi 291390734	0.22	PREDICTED: cytochrome b-c1 complex subunit 2, mitochondrial [Oryctolagus cuniculus]
gi 1256242	0.22	translocon associated protein alpha [Oryctolagus cuniculus]
gi 291389971	0.22	PREDICTED: keratin, type I cytoskeletal 18 [Oryctolagus cuniculus]
gi 291395187	0.21	PREDICTED: tubulin alpha-1B chain-like [Oryctolagus cuniculus]
gi 655730654	0.21	PREDICTED: LOW QUALITY PROTEIN: keratin, type II cytoskeletal 2 epidermal [Oryctolagus cuniculus]
gi 1551	0.21	elongation factor 1 alpha [Oryctolagus cuniculus]
gi 291395799	0.20	PREDICTED: heat shock 70 kDa protein 1B [Oryctolagus cuniculus]
gi 291388413	0.20	PREDICTED: eukaryotic translation initiation factor 3 subunit E [Oryctolagus cuniculus]
gi 291384874	0.20	PREDICTED: estradiol 17-beta-dehydrogenase 12 [Oryctolagus cuniculus]
gi 14701788	0.20	proteolipid protein 2 [Oryctolagus cuniculus]
gi 655855843	0.19	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC100358804 [Oryctolagus cuniculus]
gi 291386789	0.19	PREDICTED: endoplasmic reticulum lectin 1 [Oryctolagus cuniculus]
gi 291409112	0.18	PREDICTED: erlin-2 [Oryctolagus cuniculus]
gi 56718852	0.18	eukaryotic translation initiation factor 3 subunit 2 beta [Oryctolagus cuniculus]
gi 291411017	0.18	PREDICTED: vitamin K epoxide reductase complex subunit 1 isoform X1 [Oryctolagus cuniculus]
gi 2644966	0.18	hnRNP-E1 protein [Oryctolagus cuniculus]

gi 655856832	0.18	PREDICTED: oligosaccharyltransferase complex subunit OSTC [Oryctolagus cuniculus]
gi 655877060	0.18	PREDICTED: magnesium transporter protein 1 [Oryctolagus cuniculus]
gi 655901920	0.18	PREDICTED: AP-2 complex subunit alpha-2 [Oryctolagus cuniculus]
gi 291387898	0.17	PREDICTED: vesicular integral-membrane protein VIP36 [Oryctolagus cuniculus]
gi 655779421	0.17	PREDICTED: LOW QUALITY PROTEIN: peptidyl-prolyl cis-trans isomerase FKBP2 [Oryctolagus cuniculus]
gi 291389217	0.16	PREDICTED: keratin, type II cytoskeletal 1 [Oryctolagus cuniculus]
gi 1703	0.16	gamma non-muscle actin [Oryctolagus cuniculus]
gi 655900781	0.16	PREDICTED: glutamine synthetase [Oryctolagus cuniculus]
gi 655901923	0.16	PREDICTED: AP-2 complex subunit alpha-2-like, partial [Oryctolagus cuniculus]
gi 291391538	0.16	PREDICTED: LOW QUALITY PROTEIN: actin, cytoplasmic 1 [Oryctolagus cuniculus]
gi 291394422	0.16	PREDICTED: signal peptidase complex catalytic subunit SEC11C [Oryctolagus cuniculus]
gi 655879413	0.16	PREDICTED: ADP-ribosylation factor-like protein 8B [Oryctolagus cuniculus]
gi 655705917	0.15	PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2 isoform X1 [Oryctolagus cuniculus]
gi 655655071	0.15	PREDICTED: clathrin interactor 1 isoform X1 [Oryctolagus cuniculus]
gi 655632699	0.15	PREDICTED: AP-3 complex subunit sigma-1 [Oryctolagus cuniculus]
gi 291414663	0.15	PREDICTED: eukaryotic translation initiation factor 3 subunit L [Oryctolagus cuniculus]
gi 655853277	0.15	PREDICTED: translocation protein SEC62 [Oryctolagus cuniculus]
gi 291386382	0.15	PREDICTED: protein RMD5 homolog A [Oryctolagus cuniculus]
gi 655851123	0.14	PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit [Oryctolagus cuniculus]
gi 291406117	0.14	PREDICTED: ras-related protein Rab-5C [Oryctolagus cuniculus]
gi 6070211	0.14	eukaryotic polypeptide chain release factor 1 [Oryctolagus cuniculus]
gi 218456212	0.14	ATP synthase subunit O, mitochondrial (predicted) [Oryctolagus cuniculus]
gi 291401874	0.14	PREDICTED: membrane-associated progesterone receptor component 2 [Oryctolagus cuniculus]
gi 512125535	0.14	Chain L, Human Tweak In Complex With The Fab Fragment Of A Neutralizing Antibody
gi 398395	0.13	elongation factor 1 beta [Oryctolagus cuniculus]
gi 291393631	0.13	PREDICTED: cytochrome b-c1 complex subunit 1, mitochondrial [Oryctolagus cuniculus]
gi 655852190	0.13	PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B [Oryctolagus cuniculus]
gi 291398127	0.13	PREDICTED: vesicle-trafficking protein SEC22b [Oryctolagus cuniculus]
gi 291404870	0.13	PREDICTED: inactive pancreatic lipase-related protein 1 [Oryctolagus cuniculus]
gi 655893116	0.13	PREDICTED: serine/arginine-rich splicing factor 2 [Oryctolagus cuniculus]
gi 291390078	0.13	PREDICTED: eukaryotic translation initiation factor 3 subunit K [Oryctolagus cuniculus]
gi 291384261	0.13	PREDICTED: signal peptidase complex subunit 2 [Oryctolagus cuniculus]
gi 655863531	0.13	PREDICTED: ER membrane protein complex subunit 7 isoform X1 [Oryctolagus cuniculus]
gi 291394621	0.12	PREDICTED: aquaporin-1 [Oryctolagus cuniculus]
gi 291401067	0.12	PREDICTED: V-type proton ATPase subunit B, brain isoform [Oryctolagus cuniculus]
gi 291385707	0.12	PREDICTED: transmembrane protein 33 isoform X1 [Oryctolagus cuniculus]
gi 129730	0.12	RecName: Full=Protein disulfide-isomerase; Short=PDI; AltName: Full=Cellular thyroid hormone-binding protein;
gi 291404901	0.12	PREDICTED: thioredoxin-dependent peroxide reductase, mitochondrial [Oryctolagus cuniculus]
gi 217030873	0.12	protein disulfide isomerase-associated 3 precursor (predicted) [Oryctolagus cuniculus]
gi 655614831	0.12	PREDICTED: serine/arginine-rich splicing factor 7 isoform X1 [Oryctolagus cuniculus]
gi 291394118	0.12	PREDICTED: NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial [Oryctolagus cuniculus]

gi 291396950	0.12	PREDICTED: dnaJ homolog subfamily B member 6 [Oryctolagus cuniculus]
gi 655845589	0.12	PREDICTED: LOW QUALITY PROTEIN: transmembrane and coiled-coil domain-containing protein 1 [Oryctolagus cuniculus]
gi 655886940	0.12	PREDICTED: ATP synthase subunit f, mitochondrial [Oryctolagus cuniculus]
gi 291389142	0.12	PREDICTED: methyltransferase-like protein 7A [Oryctolagus cuniculus]
gi 291391293	0.12	PREDICTED: synaptophysin-like protein 1 [Oryctolagus cuniculus]
gi 291390349	0.11	PREDICTED: enhancer of mRNA-decapping protein 4 isoform X2 [Oryctolagus cuniculus]
gi 291389209	0.11	PREDICTED: keratin, type II cytoskeletal 5 isoform X1 [Oryctolagus cuniculus]
gi 655805242	0.11	PREDICTED: nodal modulator 1 [Oryctolagus cuniculus]
gi 655891744	0.11	PREDICTED: AP-3 complex subunit beta-1 [Oryctolagus cuniculus]
gi 291388238	0.11	PREDICTED: copine-3 [Oryctolagus cuniculus]
gi 291384976	0.11	PREDICTED: NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial [Oryctolagus cuniculus]
gi 291407221	0.11	PREDICTED: peroxiredoxin-4 isoform X3 [Oryctolagus cuniculus]
gi 655890232	0.11	PREDICTED: ER membrane protein complex subunit 3 [Oryctolagus cuniculus]
gi 655897861	0.11	PREDICTED: LOW QUALITY PROTEIN: toll-interacting protein [Oryctolagus cuniculus]
gi 291383611	0.10	PREDICTED: signal recognition particle receptor subunit alpha [Oryctolagus cuniculus]
gi 291391168	0.10	PREDICTED: staphylococcal nuclease domain-containing protein 1 [Oryctolagus cuniculus]
gi 291410136	0.10	PREDICTED: calnexin isoform X3 [Oryctolagus cuniculus]
gi 6653663	0.10	voltage-dependent anion channel 2 [Oryctolagus cuniculus]
gi 655854861	0.10	PREDICTED: V-type proton ATPase catalytic subunit A [Oryctolagus cuniculus]
gi 291403258	0.10	PREDICTED: thrombospondin-1 [Oryctolagus cuniculus]
gi 291413344	0.10	PREDICTED: alpha-soluble NSF attachment protein [Oryctolagus cuniculus]
gi 655840030	0.10	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC100343399 [Oryctolagus cuniculus]
gi 291392793	0.10	PREDICTED: prohibitin-2 [Oryctolagus cuniculus]
gi 291388415	0.09	PREDICTED: ER membrane protein complex subunit 2 [Oryctolagus cuniculus]
gi 676873107	0.09	immunoglobulin gamma heavy chain constant region, partial [Oryctolagus cuniculus cuniculus]
gi 655894769	0.09	PREDICTED: dnaJ homolog subfamily B member 1 [Oryctolagus cuniculus]
gi 655865611	0.09	PREDICTED: eukaryotic initiation factor 4A-I-like [Oryctolagus cuniculus]
gi 291409601	0.08	PREDICTED: DNA damage-binding protein 1 [Oryctolagus cuniculus]
gi 291410132	0.08	PREDICTED: RUN and FYVE domain-containing protein 1 [Oryctolagus cuniculus]
gi 291388434	0.08	PREDICTED: eukaryotic translation initiation factor 3 subunit H [Oryctolagus cuniculus]
gi 655606040	0.08	PREDICTED: mitochondrial inner membrane protein isoform X1 [Oryctolagus cuniculus]
gi 655882251	0.08	PREDICTED: serine/arginine-rich splicing factor 6 [Oryctolagus cuniculus]
gi 655601210	0.08	PREDICTED: etoposide-induced protein 2.4 homolog [Oryctolagus cuniculus]
gi 291388107	0.08	PREDICTED: translocating chain-associated membrane protein 1 [Oryctolagus cuniculus]
gi 655700678	0.08	PREDICTED: LOW QUALITY PROTEIN: protein LYRIC [Oryctolagus cuniculus]
gi 655887896	0.08	PREDICTED: secretory carrier-associated membrane protein 2 [Oryctolagus cuniculus]
gi 655900241	0.07	PREDICTED: LOW QUALITY PROTEIN: probable cation-transporting ATPase 13A1 [Oryctolagus cuniculus]
gi 655829498	0.07	PREDICTED: LOW QUALITY PROTEIN: ATP-binding cassette sub-family B member 6, mitochondrial [Oryctolagus cuniculus]
gi 655899323	0.07	PREDICTED: LOW QUALITY PROTEIN: cysteine--tRNA ligase, cytoplasmic-like [Oryctolagus cuniculus]
gi 655890200	0.07	PREDICTED: ER degradation-enhancing alpha-mannosidase-like protein 3 [Oryctolagus cuniculus]
gi 655889921	0.07	PREDICTED: protein disulfide-isomerase A6 [Oryctolagus cuniculus]

gi 291406131	0.07	PREDICTED: V-type proton ATPase 116 kDa subunit a isoform 1 isoform X3 [Oryctolagus cuniculus]
gi 291402056	0.07	PREDICTED: saccharopine dehydrogenase-like oxidoreductase [Oryctolagus cuniculus]
gi 217916954	0.07	dyskerin (predicted), partial [Oryctolagus cuniculus]
gi 291412860	0.06	PREDICTED: fatty aldehyde dehydrogenase isoform X1 [Oryctolagus cuniculus]
gi 291394426	0.06	PREDICTED: protein ERGIC-53 isoform X1 [Oryctolagus cuniculus]
gi 655899696	0.06	PREDICTED: AP-2 complex subunit alpha-1 [Oryctolagus cuniculus]
gi 655845860	0.06	PREDICTED: DDB1- and CUL4-associated factor 8 isoform X1 [Oryctolagus cuniculus]
gi 655827020	0.06	PREDICTED: metalloredoxase STEAP3 isoform X1 [Oryctolagus cuniculus]
gi 655892765	0.06	PREDICTED: E3 ubiquitin-protein ligase makorin-1 [Oryctolagus cuniculus]
gi 655605354	0.06	PREDICTED: serine/threonine-protein kinase MARK2-like [Oryctolagus cuniculus]
gi 655898441	0.06	PREDICTED: sorting nexin-8 [Oryctolagus cuniculus]
gi 655896277	0.06	PREDICTED: tubulin beta-3 chain [Oryctolagus cuniculus]
gi 655833400	0.06	PREDICTED: protein transport protein Sec61 subunit alpha isoform 1 [Oryctolagus cuniculus]
gi 291393196	0.05	PREDICTED: tripeptidyl-peptidase 2 isoform X1 [Oryctolagus cuniculus]
gi 655790980	0.05	PREDICTED: V-type proton ATPase subunit d 1 [Oryctolagus cuniculus]
gi 655865898	0.05	PREDICTED: glutamate dehydrogenase 1, mitochondrial [Oryctolagus cuniculus]
gi 655881614	0.05	PREDICTED: protein OS-9 [Oryctolagus cuniculus]
gi 655864474	0.05	PREDICTED: signal recognition particle 54 kDa protein isoform X3 [Oryctolagus cuniculus]
gi 291401168	0.05	PREDICTED: ATP-binding cassette sub-family E member 1 [Oryctolagus cuniculus]
gi 655839976	0.04	PREDICTED: phosphatidylinositol-glycan-specific phospholipase D isoform X1 [Oryctolagus cuniculus]
gi 2581793	0.04	glucose-regulated protein GRP94, partial [Oryctolagus cuniculus]
gi 655841556	0.04	PREDICTED: heat shock protein HSP 90-beta isoform X1 [Oryctolagus cuniculus]
gi 291392087	0.04	PREDICTED: NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial [Oryctolagus cuniculus]
gi 655814551	0.04	PREDICTED: protein disulfide-isomerase A4 [Oryctolagus cuniculus]
gi 2384762	0.04	glycogen synthase [Oryctolagus cuniculus]
gi 655676341	0.03	PREDICTED: aspartyl/asparaginyl beta-hydroxylase [Oryctolagus cuniculus]
gi 655869173	0.03	PREDICTED: sarcoplasmic/endoplasmic reticulum calcium ATPase 3 isoform X1 [Oryctolagus cuniculus]
gi 291403822	0.03	PREDICTED: nuclear export mediator factor NEMF isoform X2 [Oryctolagus cuniculus]
gi 5805341	0.03	glycoprotein IIb [Oryctolagus cuniculus]
gi 655851212	0.03	PREDICTED: ER membrane protein complex subunit 1 isoform X4 [Oryctolagus cuniculus]
gi 655870955	0.03	PREDICTED: importin subunit beta-1 [Oryctolagus cuniculus]
gi 291387698	0.03	PREDICTED: Ia-related protein 1 [Oryctolagus cuniculus]
gi 291389423	0.03	PREDICTED: extended synaptotagmin-1 [Oryctolagus cuniculus]
gi 655873131	0.03	PREDICTED: C-1-tetrahydrofolate synthase, cytoplasmic [Oryctolagus cuniculus]
gi 655873297	0.03	PREDICTED: protein sel-1 homolog 1 [Oryctolagus cuniculus]
gi 291385746	0.03	PREDICTED: NF-X1-type zinc finger protein NFXL1 [Oryctolagus cuniculus]
gi 291405608	0.03	PREDICTED: AP-2 complex subunit beta isoform X1 [Oryctolagus cuniculus]
gi 291408948	0.02	PREDICTED: fibronectin type-III domain-containing protein 3A isoform X1 [Oryctolagus cuniculus]
gi 291414790	0.02	PREDICTED: vigilin [Oryctolagus cuniculus]
gi 655848769	0.02	PREDICTED: glutamate-rich protein 3 isoform X1 [Oryctolagus cuniculus]
gi 291404146	0.02	PREDICTED: zinc finger SWIM domain-containing protein 8 isoform X4 [Oryctolagus cuniculus]
gi 655875291	0.02	PREDICTED: DNA polymerase alpha catalytic subunit [Oryctolagus cuniculus]

gi 655862559	0.01	PREDICTED: vacuolar protein sorting-associated protein 13C [Oryctolagus cuniculus]
--------------	------	--

List of Abbreviations

Å (unit)	Ångstroem
ΔG	Gibbs free energy
ΔH	Change in enthalpy
ΔS	Change in entropy
μg	Microgram
μL	Microliter
μM	Micromolar
3D	Three-dimensional
aa	Amino acid
aa-tRNA	Aminoacyl-transfer ribonucleic acid
<i>A. fulgidus</i>	Archeoglobus fulgidus
AgIB	Archaeal glycosylation B
A-site	Aminoacyl-site
CET	Cryo-electron tomography
CHX	Cycloheximide
<i>C. lari</i>	Camphylobacter lari
C-terminal	Carboxyterminal
CTF	Contrast transfer function
CMV	Cytomegalovirus
cryo-EM	Cryo-electron microscopy
DAD1	Defender against apoptotic cell-death 1
DC	Decoding center
DNA	Deoxyribonucleic acid
Dol	Dolichol
DTT	Dithiothreitol
ECL	Electro chemiluminescence
<i>E. coli</i>	Escherichia coli
eEF	Eukaryotic elongation factor

eIF	Eukaryotic initiation factor
EL	External loop
EMC	ER membrane protein complex
emPAI	Exponentially modified protein abundance index
Endo H	Endoglycosidase H
eq (unit)	Equivalent
ER	Endoplasmic reticulum
eRF1	Eukaryotic release factor 1
E-site	Exit-site
F	Free (non-membrane bound)
FSC	Fourier shell correlation
GIFT domain	Gliding/intraflagellar transport domain
GPCR	G-protein coupled receptor
GTP	Guanosine triphosphate
<i>H. sapiens</i>	Homo sapiens
HA	Hemagglutinin
HRP	Horseradish peroxidase
IAP	Implantation-associated protein 1
Kan	Kanamycin
KCP2	Keratinocyte associated protein 2
kDa	Kilodalton
L	Loop
LLO	Lipid-linked oligosaccharide
LSU	Large ribosomal subunit
MagT1	Magnesium transporter 1
MB	Membrane bound
mM	Millimolar
mRNA	Messenger RNA
MS	Mass spectrometry
MW	Molecular weight
NC	Nascent chain
ng	nanogram
N-glycosylation	Asparagine-linked glycosylation
N-terminal	Aminoterminal
OD	Optical density
OP	Opsin

ORF	Open reading frame
OST	Oligosaccharyltransferase
OST48	Oligosaccharyltransferase subunit of 48 kilodalton
PAA	Poly-acrylamide
PAGE	Polyacrylamide gel electrophoresis
PAT-10	Protein associated with the ER translocon of 10 kilodalton
PCC	Protein conducting channel
PCR	Polymerase chain reaction
pdb	Protein data bank
PglB	Protein glycosylation B
PKRM	Puromycin/High-salt treated rough microsomal membranes
PP_i	Pyrophosphate
P-site	Peptidyl-site
PTC	Peptidyl transferase center
RF	Release factor
RM	Rough microsomal membranes
RT	Room temperature
RNC	Ribosome-nascent chain complex
rpm	Revolutions per minute
RPN1	Ribophorin 1
RPN2	Ribophorin 2
RRL	Rabbit reticulocyte lysate
rRNA	Ribosomal ribonucleic acid
RTC	Ribosome-translocon complex
S (unit)	Svedberg unit
SA	Signal anchor
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SD	Shine-Dalgarno
SDS	Sodium dodecyl sulfate
SP	Signal peptide
SR	Signal recognition particle receptor
SRP	Signal recognition particle
SS	Signal sequence
SSR	Signal sequence receptor complex (also: TRAP)
SSU	Small ribosomal subunit
STT3	Staurosporin and temperature sensitive mutant 3

TC	Ternary complex
TCA	Trichloroacetic acid
TM	Transmembrane
TMD	Transmembrane domain
TMEM258	Transmembrane protein 258
TMP	Transmembrane protein
TMS	Transmembrane segment
TRAM	Translocating chain-associating membrane protein
TRAP	Translocon-associated protein complex
tRNA	Transfer ribonucleic acid
TUSC3	Tumor suppressor candidate 3
U	units
UTR	Untranslated region
WB	Western blot
WT	Wild type

References

- Afonine, P. V. et al. (2012). "Towards automated crystallographic structure refinement with phenix.refine". In: *Acta Crystallographica Section D: Biological Crystallography* 68, pp. 352–367.
- Akopian, D. et al. (2013). "Signal Recognition Particle: An Essential Protein-Targeting Machine". In: *Annual Review of Biochemistry* 82, pp. 693–721.
- Almeida, P. F. et al. (2012). "Hydrogen-bond energetics drive helix formation in membrane interfaces". In: *Biochimica et Biophysica Acta - Biomembranes* 1818, pp. 178–182.
- Anger, A. et al. (2013). "Structures of the human and Drosophila 80S ribosome". In: *Nature* 497, pp. 80–87.
- Bai, L. et al. (2018). "The atomic structure of a eukaryotic oligosaccharyltransferase complex". In: *Nature* 555, pp. 328–333.
- Ban, N. et al. (2000). "The complete atomic structure of the large ribosomal subunit at 2.4 Å Resolution". In: *Science* 289, pp. 905–920.
- Ban, N. et al. (2014). "A new system for naming ribosomal proteins". In: *Current Opinion in Structural Biology* 24, pp. 165–169.
- Beatson, S. and C. P. Ponting (2004). "GIFT domains: Linking eukaryotic intraflagellar transport and glycosylation to bacterial gliding". In: *Trends in Biochemical Sciences* 29, pp. 396–399.
- Becker, T. et al. (2009). "Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome". In: *Science* 326, pp. 1369–1373.
- Beckmann, R. et al. (2001). "Architecture of the protein-conducting channel associated with the translating 80S ribosome". In: *Cell* 107, pp. 361–372.
- Ben-Shem, A. et al. (2011). "The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution". In: *Science* 334, pp. 1524–1529.
- Bhushan, S. et al. (2010). "Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide". In: *Molecular Cell* 40, pp. 138–146.
- Bischoff, L. et al. (2014). "Visualization of a polytopic membrane protein during SecY-mediated membrane insertion". In: *Nature Communications* 5, p. 4103.

- Blobel, G. (1980). "Intracellular protein topogenesis". In: *Proceedings of the National Academy of Sciences* 77, pp. 1496–1500.
- Blobel, G. and B. Dobberstein (1975). "Transfer of Proteins across Membranes II . Reconstitution of Functional Rough Microsomes from Heterologous Components". In: *The Journal of Cell Biology* 67, pp. 852–862.
- Blomen, V. A. et al. (2015). "Gene essentiality and synthetic lethality in haploid human cells". In: *Science* 350, pp. 1092–1096.
- Borel, A. C. and S. M. Simon (1996). "Biogenesis of polytopic membrane proteins: Membrane segments assemble within translocation channels prior to membrane integration". In: *Cell* 85, pp. 379–389.
- Braunger, K. et al. (2018). "Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum". In: *Science* 360, pp. 215–219.
- Brown, A. et al. (2015). "Structural basis for stop codon recognition in eukaryotes". In: *Nature* 524, pp. 493–496.
- Caramelo, J. J. and A. J. Parodi (2015). "A sweet code for glycoprotein folding". In: *FEBS Letters* 589, pp. 3379–3387.
- Cherepanova, N. et al. (2016). "N-linked glycosylation and homeostasis of the endoplasmic reticulum". In: *Current Opinion in Cell Biology* 41, pp. 57–65.
- Cherepanova, N. A. and R. Gilmore (2016). "Mammalian cells lacking either the cotranslational or posttranslocational oligosaccharyltransferase complex display substrate-dependent defects in asparagine linked glycosylation". In: *Scientific Reports* 6, pp. 1–12.
- Cherepanova, N. A. et al. (2014). "Oxidoreductase activity is necessary for N-glycosylation of cysteine-proximal acceptor sites in glycoproteins". In: *The Journal of Cell Biology* 206, pp. 525–539.
- Christianson, J. C. et al. (2012). "Defining human ERAD networks through an integrative mapping strategy". In: *Nature Cell Biology* 14, pp. 93–105.
- Crimaudo, C et al. (1987). "Human ribophorins I and II: the primary structure and membrane topology of two highly conserved rough endoplasmic reticulum-specific glycoproteins". In: *EMBO Journal* 6, pp. 75–82.
- Cross, B. C. S. and S. High (2009). "Dissecting the physiological role of selective transmembrane-segment retention at the ER translocon". In: *Journal of Cell Science* 122, pp. 1768–1777.
- Cymer, F. et al. (2015). "Mechanisms of Integral Membrane Protein Insertion and Folding". In: *Journal of Molecular Biology* 427, pp. 999–1022.

- De Marothy, M. T. and A. Elofsson (2015). "Marginally hydrophobic transmembrane α -helices shaping membrane protein folding". In: *Protein Science* 24, pp. 1057–1074.
- Degnin, C. R. et al. (1993). "Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript." In: *Journal of virology* 67, pp. 5514–5521.
- Dell, A. et al. (2010). "Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes". In: *International Journal of Microbiology* 2010.
- Devaraneni, P. K. et al. (2011). "Stepwise insertion and inversion of a type II signal anchor sequence in the ribosome-Sec61 translocon complex". In: *Cell* 146, pp. 134–147.
- Drozdetskiy, A. et al. (2015). "JPred4: A protein secondary structure prediction server". In: *Nucleic Acids Research* 43, W389–W394.
- Dumax-Vorzet, A. et al. (2013). "OST4 is a subunit of the mammalian oligosaccharyltransferase required for efficient N-glycosylation". In: *Journal of Cell Science* 126, pp. 2595–2606.
- Egea, P. F. and R. M. Stroud (2010). "Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes." In: *Proceedings of the National Academy of Sciences of the United States of America* 107, pp. 17182–17187.
- Emsley, P. and K. Cowtan (2004). "Coot: Model-building tools for molecular graphics". In: *Acta Crystallographica Section D: Biological Crystallography* 60, pp. 2126–2132.
- Fetrow, J. S. et al. (2001). "Genomic-scale comparison of sequence- and structure-based methods of function prediction: does structure provide additional insight?" In: *Protein Science* 10, pp. 1005–1014.
- Fu, J. et al. (1997). "Interactions among subunits of the oligosaccharyltransferase complex". In: *Journal of Biological Chemistry* 272, pp. 29687–29692.
- Gavel, Y. and G. V. Heijne (1990). "Sequence differences between glycosylated and non-glycosylated asn-x-thr/ser acceptor sites: Implications for protein engineering". In: *Protein Engineering, Design and Selection* 3, pp. 433–442.
- Gayen, S. and C. Kang (2011). "Solution structure of a human minimembrane protein Ost4, a subunit of the oligosaccharyltransferase complex". In: *Biochemical and Biophysical Research Communications* 409, pp. 572–576.
- Goder, V. and M. Spiess (2003). "Molecular mechanism of signal sequence orientation in the endoplasmic reticulum". In: *EMBO Journal* 22, pp. 3645–3653.

- Gogala, M. et al. (2014). "Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion." In: *Nature* 506, pp. 107–10.
- Goldman, B. M. and G. Blobel (1981). "In Vitro Biosynthesis, Core Glycosylation, and Membrane Integration of Opsin". In: *The Journal of Cell Biology* 90, pp. 236–242.
- Görllich, D. and T. a. Rapoport (1993). "Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane". In: *Cell* 75, pp. 615–630.
- Görllich, D. et al. (1990). "The signal sequence receptor has a second subunit and is part of a translocation complex in the endoplasmic reticulum as probed by bifunctional reagents". In: *The Journal of Cell Biology* 111, pp. 2283–2294.
- Görllich, D et al. (1992). "A protein of the endoplasmic reticulum involved early in polypeptide translocation." In: *Nature* 357, pp. 47–52.
- Hanover, J. A. and W. J. Lennarz (1980). "N-Linked glycoprotein assembly. Evidence that oligosaccharide attachment occurs within the lumen of the endoplasmic reticulum." In: *Journal of Biological Chemistry* 255, pp. 3600–3604.
- Hargrave, P. A. et al. (1983). "The Structure of Bovine Rhodopsin". In: *Biophysics of Structure and Mechanism* 9, pp. 235–244.
- Hargrave, P. A. (1977). "The amino-terminal tryptic peptide of bovine rhodopsin A glycopeptide containing two sites of oligosaccharide attachment". In: *Biochimica et Biophysica Acta* 492, pp. 83–94.
- Hartmann, E et al. (1993). "A tetrameric complex of membrane proteins in the endoplasmic reticulum." In: *European Journal of Biochemistry* 214, pp. 375–381.
- Hedin, L. E. et al. (2010). "Membrane Insertion of Marginally Hydrophobic Transmembrane Helices Depends on Sequence Context". In: *Journal of Molecular Biology* 396, pp. 221–229.
- Heinrich, S. U. and T. A. Rapoport (2003). "Cooperation of transmembrane segments during the integration of a double-spanning protein into the ER membrane". In: *EMBO Journal* 22, pp. 3654–3663.
- Heinrich, S. U. et al. (2000). "The Sec61p Complex Mediates the Integration of a Membrane Protein by Allowing Lipid Partitioning of the Transmembrane Domain". In: *Cell* 102, pp. 233–244.
- Helenius, A. and M. Aeby (2004). "Roles of N-linked glycans in the endoplasmic reticulum". In: *Annual Review of Biochemistry* 73, pp. 1019–49.
- Helenius, J. et al. (2002). "Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein". In: *Nature* 415, pp. 447–450.

- Helmers, J. et al. (2003). "The β -subunit of the protein-conducting channel of the endoplasmic reticulum functions as the guanine nucleotide exchange factor for the β -subunit of the signal recognition particle receptor". In: *The Journal of Biological Chemistry* 278, pp. 23686–23690.
- Hennet, T. and J. Cabalzar (2015). "Congenital disorders of glycosylation: A concise chart of glyocalyx dysfunction". In: *Trends in Biochemical Sciences* 40, pp. 377–384.
- Hessa, T. et al. (2005). "Recognition of transmembrane helices by the endoplasmic reticulum translocon." In: *Nature* 433, pp. 377–381.
- Hessa, T. et al. (2007). "Molecular code for transmembrane-helix recognition by the Sec61 translocon." In: *Nature* 450, pp. 1026–1030.
- High, S. et al. (1993). "Site-specific photocross-linking reveals that Sec61p and TRAM contact different regions of a membrane-inserted signal sequence." In: *The Journal of Biological Chemistry* 268, pp. 26745–26751.
- Higy, M. et al. (2004). "Topogenesis of membrane proteins at the endoplasmic reticulum". In: *Biochemistry* 43, pp. 12716–12722.
- Igura, M. et al. (2008). "Structure-guided identification of a new catalytic motif of oligosaccharyltransferase." In: *The EMBO Journal* 27, pp. 234–43.
- Ishihama, Y. et al. (2005). "Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein". In: *Molecular & Cellular Proteomics* 4, pp. 1265–1272.
- Ismail, N. et al. (2006). "Active and passive displacement of transmembrane domains both occur during opsin biogenesis at the Sec61 translocon." In: *Journal of Cell Science* 119, pp. 2826–2836.
- Ismail, N. et al. (2008). "Specific transmembrane segments are selectively delayed at the ER translocon during opsin biogenesis." In: *The Biochemical Journal* 411, pp. 495–506.
- Jiang, Y. et al. (2008). "An interaction between the SRP receptor and the translocon is critical during cotranslational protein translocation". In: *The Journal of Cell Biology* 180, pp. 1149–1161.
- Jones, M. B. et al. (2009). "Structure and synthesis of polyisoprenoids used in N-glycosylation across the three domains of life". In: *Biochimica et Biophysica Acta - General Subjects* 1790, pp. 485–494.
- Jonikas, M. C. et al. (2009). "Comprehensive Characterization of Genes Required for Protein Folding in the Endoplasmic Reticulum". In: *Science* 323, pp. 1693–1697.

- Junne, T. and M. Spiess (2017). "Integration of transmembrane domains is regulated by their downstream sequences". In: *The Journal of Cell Science* 130, pp. 372–381.
- Kalies, K. U. et al. (1998). "The β subunit of the Sec61 complex facilitates cotranslational protein transport and interacts with the signal peptidase during translocation". In: *The Journal of Cell Biology* 141, pp. 887–894.
- Kanki, T. et al. (2002). "The tenth membrane region of band 3 is initially exposed to the luminal side of the endoplasmic reticulum and then integrated into a partially folded band 3 intermediate". In: *Biochemistry* 41, pp. 13973–13981.
- Karaoglu, D. et al. (1995). "Functional characterization of Ost3p. Loss of the 34-kD subunit of the *Saccharomyces cerevisiae* oligosaccharyltransferase results in biased underglycosylation of acceptor substrates". In: *Journal of Cell Biology* 130, pp. 567–577.
- Karaoglu, D. et al. (1997). "The highly conserved Stt3 protein is a subunit of the yeast oligosaccharyltransferase and forms a subcomplex with Ost3p and Ost4p". In: *Journal of Biological Chemistry* 272, pp. 32513–32520.
- Kauko, A. et al. (2010). "Repositioning of Transmembrane α -Helices during Membrane Protein Folding". In: *Journal of Molecular Biology* 397, pp. 190–201.
- Kelkar, A. and B. Dobberstein (2009). "Sec61 β , a subunit of the Sec61 protein translocation channel at the Endoplasmic Reticulum, is involved in the transport of Gurken to the plasma membrane". In: *BMC Cell Biology* 10, pp. 1–14.
- Kelleher, D. J. and R. Gilmore (2006). "An evolving view of the eukaryotic oligosaccharyltransferase". In: *Glycobiology* 16, pp. 47–62.
- Kelleher, D. J. et al. (1992). "Oligosaccharyltransferase activity is associated with a protein complex composed of ribophorins I and II and a 48 kd protein". In: *Cell* 69, pp. 55–65.
- Kelleher, D. J. and R. Gilmore (1997). "DAD1, the defender against apoptotic cell death, is a subunit of the mammalian oligosaccharyltransferase." In: *Proceedings of the National Academy of Sciences of the United States of America* 94, pp. 4994–4999.
- Kelleher, D. J. et al. (2003). "Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties". In: *Molecular Cell* 12, pp. 101–111.
- Kelly, L. et al. (2015). "The Phyre2 web portal for protein modelling, prediction, and analysis". In: *Nature Protocols* 10, pp. 845–858.
- Khatter, H. et al. (2015). "Structure of the human 80S ribosome". In: *Nature* 520, pp. 640–645.

- Kihara, A et al. (1995). "FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit." In: *Proceedings of the National Academy of Sciences of the United States of America* 92, pp. 4532–4536.
- Kim, H. et al. (2003). "Determination of the membrane topology of Ost4p and its subunit interactions in the oligosaccharyltransferase complex in *Saccharomyces cerevisiae*". In: *Proceedings of the National Academy of Sciences* 100, pp. 7460–7464.
- Kim, H. et al. (2005). "Membrane topology of the STT3 subunit of the oligosaccharyl transferase complex". In: *Journal of Biological Chemistry* 280, pp. 20261–20267.
- Kimanius, D. et al. (2016). "Accelerated cryo-EM structure determination with parallelisation using GPUS in RELION-2". In: *eLife* 5, pp. 1–21.
- Kocik, L. et al. (2012). "Orientation of internal signal-anchor sequences at the sec61 translocon". In: *Journal of Molecular Biology* 424, pp. 368–378.
- Kontinen, V. P. et al. (1996). "Roles of the conserved cytoplasmic region and non-conserved carboxy-terminal region of SecE in *Escherichia coli* protein translocasen". In: *Journal of Biochemistry* 119, pp. 1124–1130.
- Kowarik, M. et al. (2006). "Definition of the bacterial N-glycosylation site consensus sequence". In: *EMBO Journal* 25, pp. 1957–1966.
- Kramer, G. et al. (2009). "The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins". In: *Nature Structural and Molecular Biology* 16, pp. 589–597.
- Kreibich, G. et al. (1978). "Proteins of rough microsomal membranes related to ribosome binding. I. Identification of ribophorins I and II, membrane proteins characteristic of rough microsomes". In: *Journal of Cell Biology* 77, pp. 464–487.
- Ladokhin, A. S. and S. H. White (1999). "Folding of amphipathic α -helices on membranes: energetics of helix formation by melittin". In: *Journal of Molecular Biology* 285, pp. 1363–1369.
- Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4". In: *Nature* 227, pp. 680–685.
- Laird, V. and S. High (1997). "Discrete cross-linking products identified during membrane protein biosynthesis". In: *Journal of Biological Chemistry* 272, pp. 1983–1989.
- Lara, P. et al. (2017). "Refined topology model of the STT3/Stt3 protein subunit of the oligosaccharyl transferase complex". In: *Journal of Biological Chemistry* 292, pp. 11349–11360.

- Lee, H. and H. Kim (2014). "Membrane topology of transmembrane proteins: determinants and experimental tools". In: *Biochemical and Biophysical Research Communications* 453, pp. 268–276.
- Li, H. et al. (2008). "Structure of the Oligosaccharyl Transferase Complex at 12 Å Resolution". In: *Structure* 16, pp. 432–440.
- Liu, B. and S.-B. Qian (2016). "Characterizing inactive ribosomes in translational profiling". In: *Translation* 4, e1138018.
- Lizak, C. et al. (2011). "X-ray structure of a bacterial oligosaccharyltransferase." In: *Nature* 474, pp. 350–355.
- Losfeld, M. E. et al. (2014). "A new congenital disorder of glycosylation caused by a mutation in SSR4, the signal sequence receptor 4 protein of the TRAP complex". In: *Human Molecular Genetics* 23, pp. 1602–1605.
- Lu, Y et al. (2000). "Reorientation of aquaporin-1 topology during maturation in the endoplasmic reticulum." In: *Molecular biology of the cell* 11, pp. 2973–2985.
- Lycklama a Nijeholt, J. A. et al. (2013). "Characterization of the supporting role of SecE in protein translocation". In: *FEBS Letters* 587, pp. 3083–3088.
- MacGrogan, D. et al. (1996). "Structure and methylation-associated silencing of a gene within a homozygously deleted region of human chromosome band 8p22". In: *Genomics* 35, pp. 55–65.
- MacKinnon, A. L. et al. (2014). "An allosteric Sec61 inhibitor traps nascent transmembrane helices at the lateral gate". In: *eLife* 2014, pp. 1–23.
- Maita, N. et al. (2010). "Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases". In: *Journal of Biological Chemistry* 285, pp. 4941–4950.
- Malaby, H. L. H. and W. R. Kobertz (2014). "The Middle X Residue Influences Co-translational N - Glycosylation Consensus Site Skipping". In: *Biochemistry* 53, pp. 4884–4893.
- Maley, F. et al. (1989). "Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases". In: *Analytical Biochemistry* 180, pp. 195–204.
- Matheisl, S. et al. (2015). "Structure of a human translation termination complex." In: *Nucleic acids research* 1, pp. 1–12.
- Matsumoto, S. et al. (2013). "Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation." In: *Proceedings of the National Academy of Sciences of the United States of America* 110, pp. 17868–17873.

- Matsumoto, S. et al. (2017). "Tethering an N-Glycosylation Sequon-Containing Peptide Creates a Catalytically Competent Oligosaccharyltransferase Complex". In: *Biochemistry* 56, pp. 602–611.
- McCormick, P. J. et al. (2003). "Cotranslational protein integration into the ER membrane is mediated by the binding of nascent chains to translocon proteins". In: *Molecular Cell* 12, pp. 329–341.
- McKenna, M. et al. (2017). "Mycolactone reveals substrate-driven complexity of Sec61-dependent transmembrane protein biogenesis". In: *Journal of Cell Science* 130, pp. 1307–1320.
- McKenna, M. et al. (2016). "Mechanistic insights into the inhibition of Sec61-dependent co- and post-translational translocation by mycolactone". In: *Journal of Cell Science* 129, pp. 1404–1415.
- Meacock, S. L. et al. (2002). "Different Transmembrane Domains Associate with Distinct Endoplasmic Reticulum Components during Membrane Integration of a Polytopic Protein". In: *Molecular Biology of the Cell* 13, pp. 4114–4129.
- Melnikov, S. et al. (2012). "One core, two shells: bacterial and eukaryotic ribosomes." In: *Nature Structural and Molecular Biology* 19, pp. 560–567.
- Ménétret, J. F. et al. (2005). "Architecture of the ribosome-channel complex derived from native membranes". In: *Journal of Molecular Biology* 348, pp. 445–457.
- Ménétret, J. F. et al. (2008). "Single Copies of Sec61 and TRAP Associate with a Non-translating Mammalian Ribosome". In: *Structure* 16, pp. 1126–1137.
- Mindell, J. A. and N. Grigorieff (2003). "Accurate determination of local defocus and specimen tilt in electron microscopy". In: *Journal of Structural Biology* 142, pp. 334–347.
- Mohorko, E. et al. (2011). "Oligosaccharyltransferase: The central enzyme of N-linked protein glycosylation". In: *Journal of Inherited Metabolic Disease* 34, pp. 869–878.
- Mothes, W. et al. (1994). "Systematic probing of the environment of a translocating secretory protein during translocation through the ER membrane." In: *EMBO Journal* 13, pp. 3973–82.
- Mothes, W. et al. (1997). "Molecular Mechanism of Membrane Protein Integration into the Endoplasmic Reticulum". In: *Cell* 89, pp. 523–533. arXiv: [339](#).
- Murphy, C. K. and J Beckwith (1994). "Residues essential for the function of SecE, a membrane component of the Escherichia coli secretion apparatus, are located in a conserved cytoplasmic region." In: *Proceedings of the National Academy of Sciences of the United States of America* 91, pp. 2557–61.

- Napiórkowska, M. et al. (2017). "Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase". In: *Nature Structural and Molecular Biology* 24, pp. 1100–1106.
- Nilsson, I. and G. Von Heijne (1993). "Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane". In: *Journal of Biological Chemistry* 268, pp. 5798–5801.
- Nilsson, I. et al. (1994). "The COOH-terminal ends of internal signal and signal-anchor sequences are positioned differently in the ER translocase". In: *Journal of Cell Biology* 126, pp. 1127–1132.
- Nyathi, Y. et al. (2013). "Co-translational targeting and translocation of proteins to the endoplasmic reticulum". In: *Biochimica et Biophysica Acta - Molecular Cell Research* 1833, pp. 2392–2402.
- Palczewski, K. (2006). "G protein-coupled receptor rhodopsin." In: *Annual Review of Biochemistry* 75, pp. 743–767.
- Palczewski, K. et al. (2000). "Crystal Structure of Rhodopsin : A G Protein – Coupled Receptor". In: *Science* 289, pp. 739–745.
- Park, E. and T. a. Rapoport (2011). "Preserving the membrane barrier for small molecules during bacterial protein translocation." In: *Nature* 473, pp. 239–242.
- Park, E. et al. (2014). "Structure of the SecY channel during initiation of protein translocation." In: *Nature* 506, pp. 102–106.
- Parnas, O. et al. (2015). "A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks". In: *Cell* 162, pp. 675–686.
- Pathak, R. et al. (1995). "Sulfhydryl Modification of the Yeast Wbp1p Inhibits Oligosaccharyl Transferase Activity". In: *Biochemistry* 34, pp. 4179–4185.
- Pech, M. et al. (2010). "Dual binding mode of the nascent polypeptide-associated complex reveals a novel universal adapter site on the ribosome". In: *Journal of Biological Chemistry* 285, pp. 19679–19687.
- Pettersen, E. F. et al. (2004). "UCSF Chimera—a visualization system for exploratory research and analysis." In: *Journal of computational chemistry* 25, pp. 1605–1612.
- Pfeffer, S. et al. (2014). "Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon." In: *Nature Communications* 5, p. 3072.
- Pfeffer, S. et al. (2015). "Structure of the native Sec61 protein-conducting channel". In: *Nature Communications* 6, p. 8403.
- Pfeffer, S. et al. (2017). "Dissecting the molecular organization of the translocon-associated protein complex". In: *Nature Communications* 8, p. 14516.

- Reiss, G. et al. (1997). "A specific screen for oligosaccharyltransferase mutations identifies the 9 kDa OST5 protein required for optimal activity in vivo and in vitro". In: *EMBO Journal* 16, pp. 1164–1172.
- Ridge, K. D. et al. (1995). "In vivo assembly of rhodopsin from expressed polypeptide fragments." In: *Proceedings of the National Academy of Sciences of the United States of America* 92, pp. 3204–3208.
- Ridge, K. D. et al. (1996). "Examining rhodopsin folding and assembly through expression of polypeptide fragments". In: *Journal of Biological Chemistry* 271, pp. 7860–7867.
- Roboti, P. and S. High (2012a). "Keratinocyte-associated protein 2 is a bona fide subunit of the mammalian oligosaccharyltransferase". In: *Journal of Cell Science* 125, pp. 220–232.
- Roboti, P. and S. High (2012b). "The oligosaccharyltransferase subunits OST48, DAD1 and KCP2 function as ubiquitous and selective modulators of mammalian N-glycosylation." In: *Journal of Cell Science* 125, pp. 3474–84.
- Rohou, A. and N. Grigorieff (2015). "CTFFIND4: Fast and accurate defocus estimation from electron micrographs". In: *Journal of Structural Biology* 192, pp. 216–221.
- Rösch, K. et al. (2000). "The topogenic contribution of uncharged amino acids on signal sequence orientation in the endoplasmic reticulum". In: *Journal of Biological Chemistry* 275, pp. 14916–14922.
- Rothman, J. E. and H. F. Lodish (1977). "Synchronised transmembrane insertion and glycosylation of a nascent membrane protein". In: *Nature* 269, pp. 775–780.
- Ruiz-Canada, C. et al. (2009). "Cotranslational and Posttranslational N-Glycosylation of Polypeptides by Distinct Mammalian OST Isoforms". In: *Cell* 136, pp. 272–283.
- Rutkowski, D. T. et al. (2001). "Substrate-specific regulation of the ribosome-translocation junction by N-terminal signal sequences". In: *Proceedings of the National Academy of Sciences of the United States of America* 98, pp. 7823–7828.
- Sanjay, A. et al. (1998). "DAD1 Is Required for the Function and the Structural Integrity of the Oligosaccharyltransferase Complex". In: *The Journal of Biological Chemistry* 273, pp. 26094–26099.
- Sato, T. et al. (2012). "STT3B-Dependent Posttranslational N-Glycosylation as a Surveillance System for Secretory Protein". In: *Molecular Cell* 47, pp. 99–110.
- Saurí, A. et al. (2007). "Sec61 α and TRAM are Sequentially Adjacent to a Nascent Viral Membrane Protein during its ER Integration". In: *Journal of Molecular Biology* 366, pp. 366–374.

- Schmeing, T. M. and V. Ramakrishnan (2009). "What recent ribosome structures have revealed about the mechanism of translation". In: *Nature* 461, pp. 1234–1242.
- Schmidt, C. et al. (2015). "Structure of the hypusinylated eukaryotic translation factor eIF-5A bound to the ribosome". In: *Nucleic Acids Research* 44, pp. 1944–1951.
- Schulz, B. L. et al. (2009). "Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency." In: *Proceedings of the National Academy of Sciences of the United States of America* 106, pp. 11061–6.
- Seppälä, S. et al. (2010). "Control of Membrane Protein Topology by a Single C-Terminal Residue". In: *Science* 328, pp. 1698–1701.
- Shaw, A. S. et al. (1988). "Evidence for the loop model of signal-sequence insertion into the endoplasmic reticulum." In: *Proceedings of the National Academy of Sciences* 85, pp. 7592–7596.
- Shibatani, T. et al. (2005). "Proteomic analysis of mammalian oligosaccharyltransferase reveals multiple subcomplexes that contain Sec61, TRAP, and two potential new subunits". In: *Biochemistry* 44, pp. 5982–5992.
- Shrimal, S. and R. Gilmore (2013). "Glycosylation of closely spaced acceptor sites in human glycoproteins". In: *Journal of Cell Science* 126, pp. 5513–5523.
- Shrimal, S. et al. (2017). "DC2 and KCP2 mediate the interaction between the oligosaccharyltransferase and the ER translocon". In: *Journal of Cell Biology* 216, pp. 3625–3638.
- Shrimal, S. et al. (2013). "Extreme C-terminal sites are posttranslocationally glycosylated by the STT3B isoform of the OST". In: *Journal of Cell Biology* 201, pp. 81–95.
- Shrimal, S. et al. (2015). "Cotranslational and posttranslocational N-glycosylation of proteins in the endoplasmic reticulum." In: *Seminars in Cell and Developmental Biology* 41, pp. 71–8.
- Sibbald, M. J. et al. (2010). "Synthetic effects of secG and secY2 mutations on exoproteome biogenesis in *Staphylococcus aureus*". In: *Journal of Bacteriology* 192, pp. 3788–3800.
- Silberstein, S. et al. (1992). "The 48-kDa subunit of the mammalian oligosaccharyltransferase complex is homologous to the essential yeast protein WBP1". In: *Journal of Biological Chemistry* 267, pp. 23658–23663.
- Simon, S. M. and G. Blobel (1991). "A protein-conducting channel in the endoplasmic reticulum". In: *Cell* 65, pp. 371–380.

- Sommer, N. et al. (2013). "TRAP assists membrane protein topogenesis at the mammalian ER membrane". In: *Biochimica et Biophysica Acta - Molecular Cell Research* 1833, pp. 3104–3111.
- Spiess, M (1995). "Heads or tails-what determines the orientation of proteins in the membrane." In: *FEBS Letters* 369, pp. 76–9.
- Spirig, U. et al. (1997). "The STT3 protein is a component of the yeast oligosaccharyltransferase complex". In: *Molecular and General Genetics* 256, pp. 628–637.
- Spirig, U. et al. (2005). "The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast". In: *Glycobiology* 15, pp. 1396–1406.
- Stanley, P. et al. (2015). "N-Glycans". In: *Essentials of Glycobiology*, pp. 1–14.
- Szymanski, C. M. et al. (1999). "Evidence for a system of general protein glycosylation in *Campylobacter jejuni*." In: *Molecular Microbiology* 32, pp. 1022–1030.
- Tamborero, S. et al. (2011). "Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM)". In: *Journal of Molecular Biology* 406, pp. 571–582.
- Tannous, A. et al. (2015). "N-linked sugar-regulated protein folding and quality control in the ER". In: *Seminars in Cell and Developmental Biology* 41, pp. 79–89.
- Te Heesen, S et al. (1992). "The yeast WBP1 is essential for oligosaccharyl transferase activity in vivo and in vitro". In: *EMBO Journal* 11, pp. 2071–2075.
- Tsirigos, K. D. et al. (2018). "Topology of membrane proteins — predictions, limitations and variations". In: *Current Opinion in Structural Biology* 50, pp. 9–17.
- Tsukazaki, T. et al. (2008). "Conformational transition of Sec machinery inferred from bacterial SecYE structures." In: *Nature* 455, pp. 988–991.
- Van den Berg, B. et al. (2004). "X-ray structure of a protein-conducting channel." In: *Nature* 427, pp. 36–44.
- Vermeire, K. et al. (2014). "Signal Peptide-Binding Drug as a Selective Inhibitor of Co-Translational Protein Translocation". In: *PLoS Biology* 12.
- Vermeire, K. et al. (2015). "Ribonuclease-neutralized pancreatic microsomal membranes from livestock for in vitro co-translational protein translocation". In: *Analytical Biochemistry* 484, pp. 102–104.
- Voigt, S. et al. (1996). "Signal sequence-dependent function of the TRAM protein during early phases of protein transport across the endoplasmic reticulum membrane". In: *Journal of Cell Biology* 134, pp. 25–35.
- Von Heijne, G (2006). "Membrane-protein topology." In: *Nature reviews. Molecular cell biology* 7, pp. 909–918.

- Von Heijne, G. (2011). "Introduction to Theme "Membrane Protein Folding and Insertion"". In: *Annual Review of Biochemistry* 80, pp. 157–160.
- Von Heijne, G. (1986). "The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology." In: *EMBO Journal* 5, pp. 3021–3027.
- Von Heijne, G. and Y. Gavel (1988). "Topogenic signals in integral membrane proteins". In: *European Journal of Biochemistry* 174, pp. 671–678.
- Voorhees, R. M. and R. S. Hegde (2016). "Structure of the Sec61 channel opened by a signal sequence". In: *Science* 351, pp. 88–91.
- Voorhees, R. M. and V. Ramakrishnan (2013). "Structural Basis of the Translational Elongation Cycle". In: *Annual Review of Biochemistry* 82, pp. 203–236.
- Voorhees, R. M. et al. (2014). "Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution". In: *Cell* 157, pp. 1632–1643.
- Voss, N. R. et al. (2006). "The Geometry of the Ribosomal Polypeptide Exit Tunnel". In: *Journal of Molecular Biology* 360, pp. 893–906.
- Walter, P. and G. Blobel (1983). "Preparation of Microsomal Membranes for Cotranslational Protein Translocation". In: *Methods in Enzymology* 96, pp. 84–93.
- Watson, H. R. et al. (2013). "Reorientation of the first signal-anchor sequence during potassium channel biogenesis at the Sec61 complex". In: *Biochemical Journal* 456, pp. 297–309.
- Welply, J. K. et al. (1983). "Substrate Recognition by Oligosaccharyltransferase - Studies on Glycosylation of Modified Asn-X-Thr/Ser Tripeptides". In: *The Journal of Biological Chemistry* 268, pp. 11856–11863.
- Whitley, P. et al. (1996). "A nascent secretory protein may traverse the ribosome/endoplasmic reticulum translocase complex as an extended chain". In: *Journal of Biological Chemistry* 271, pp. 6241–6244.
- Wiedmann, M. et al. (1987). "Direct probing of the interaction between the signal sequence of nascent preprolactin and the signal recognition particle by specific cross-linking." In: *The Journal of Cell Biology* 104, pp. 201–208.
- Wild, R. et al. (2018). "Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation". In: *Science* 550, pp. 1–12.
- Wilson, C. M. and S. High (2007). "Ribophorin I acts as a substrate-specific facilitator of N-glycosylation". In: *Journal of Cell Science* 120, pp. 648–657.
- Wilson, C. M. et al. (2005). "Ribophorin I associates with a subset of membrane proteins after their integration at the Sec61 translocon". In: *Journal of Biological Chemistry* 280, pp. 4195–4206.

- Wilson, C. M. et al. (2008). "Ribophorin I regulates substrate delivery to the oligosaccharyltransferase core." In: *Proceedings of the National Academy of Sciences of the United States of America* 105, pp. 9534–9.
- Wilson, D. N. and J. H. D. Cate (2012). "The Structure and Function of the Eukaryotic Ribosome". In: *Cold Spring Harbor Perspectives in Biology* 4, a011536.
- Wimberly, B. T. et al. (2000). "Structure of the 30S ribosomal subunit". In: *Nature* 407, pp. 327–339.
- Yu, Y. et al. (1990). "Antiribophorin antibodies inhibit the targeting to the ER membrane of ribosomes containing nascent secretory polypeptides". In: *Journal of Cell Biology* 111, pp. 1335–1342.
- Yusupova, G. and M. Yusupov (2014). "High-Resolution Structure of the Eukaryotic 80S Ribosome". In: *Annual Review of Biochemistry* 83, pp. 467–486.
- Zhang, K. (2016). "Gctf: Real-time CTF determination and correction". In: *Journal of Structural Biology* 193, pp. 1–12.
- Zheng, S. Q. et al. (2017). "MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy". In: *Nature Methods* 14, pp. 331–332.
- Zielinska, D. F. et al. (2010). "Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints". In: *Cell* 141, pp. 897–907.
- Zielinska, D. F. et al. (2012). "Mapping N-Glycosylation Sites across Seven Evolutionarily Distant Species Reveals a Divergent Substrate Proteome Despite a Common Core Machinery". In: *Molecular Cell* 46, pp. 542–548.
- Zimmer, J. et al. (2008). "Structure of a complex of the ATPase SecA and the protein-translocation channel." In: *Nature* 455, pp. 936–943.
- Zivanov, J. et al. (2018). "RELION-3: new tools for automated high-resolution cryo-EM structure determination". In: *bioRxiv*, p. 421123.
- Zubkov, S. et al. (2004). "Structural basis for the function of a minimembrane protein subunit of yeast oligosaccharyltransferase." In: *Proceedings of the National Academy of Sciences of the United States of America* 101, pp. 3821–3826.

Acknowledgement

The past four and a half years of my Masters' thesis and PhD work in the Beckmann lab have been an amazing journey.

I would like to thank Prof. Roland Beckmann for offering me a PhD position in his group. You have been an inspiration, both scientifically and personally. You really encourage people to develop their own style of research and provide excellent guidance on the way to do so.

Thank you to Prof. Karl-Peter Hopfner for taking the time to be my second thesis referee and also for shaping the GRK1721 graduate school the way that you do. It is a fantastic program which has enabled me to build a great network of Munich PhD students in structural biology and I have always enjoyed the GRK1721 events.

I would also like to thank PD Dr. Gregor Witte, Prof. Friedrich Förster and Prof. Klaus Förstemann for their expertise and valuable advice in my thesis advisory committee and Prof. Alexej Kedrov for being part of my defense committee.

I am very grateful to Prof. Stefan Pfeffer for being a fantastic collaborator and co-author. You really made the difference when things were (almost) getting ahead of us.

Furthermore, many thanks goes to the Boehringer Ingelheim Fonds, for supporting my thesis work way beyond a monthly stipend. The BIF events have always been and continue to be an inspiring combination of science and the people behind it.

Thank you to everyone in the Beckmann lab for making this place my second home. No matter what science had in stock for me, I always enjoyed coming to work and that's because you guys rock!

On that note, a few people deserve special mentioning: Birgitta and Thomas - thank you for being Roland's right and left hand - not only when he's gone 'Surfin' USA'. Joanna, Andrea and Heidi - I am convinced that the lab would fall apart in less than a month if it weren't for you! I know that I will miss many of the luxuries you spoil us with at my future workplace. 'The IT guys' - your patience, skill and good humor are incredible... The EM facility - thank you for taking the responsibility for our samples and transforming it into the coolest data ever ;-). Chris - you've been an awesome Maxi-Me and I couldn't say it served me any bad. André - supervisor

at first but colleague and friend soon enough. Thank you for all talks about work and the world. Anne - thank you for discussing all my stupid or not-so-stupid questions and for being the best partner-in-crime on worst-ever-puns I could have hoped for. You always brightened up my days. And last but not least, Hanna - I have no words. The best things in life come for free. These people and many others made the Gene Center a place which I will always keep in good memory. Finally, I would like to thank my friends and my family for supporting me in countless ways on my journey to this degree. It means a lot!

Let us read, and let us dance; these two amusements will never do any harm to the world.
- Voltaire